DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

High-level neural structures constrain visual behavior

Citation

Cohen, Michael A. 2014. High-level neural structures constrain visual behavior. Doctoral dissertation, Harvard University.

Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:12269824

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

High-level neural structures constrain visual behavior

A dissertation presented
By
Michael A. Cohen
To
The Department of Psychology

In partial fulfillment of the requirements
For the degree of
Doctor of Philosophy
In the subject of
Psychology

Harvard University
Cambridge, Massachusetts
April, 2014

©2014 – Michael A. Cohen
All rights reserved.

Dissertation advisor: Michael A. Cohen
Professors Ken Nakayama & George Alvarez

High-level neural structures constrain visual behavior

Abstract

Visual cognition is notoriously limited: only a finite amount of information can be fully processed at a given instant. What is the source of these limitations? Here, we suggest that the organization of higher-level visual cortex into content-specific channels constrains information processing across the visual system. Each channel is primarily involved in representing one particular type of visual content (e.g. faces, cars, certain types of shapes, etc.). Furthermore, each channel has a finite processing capacity/bandwidth and is limited in the amount of information it can process. When multiple items are simultaneously presented across space, or quickly in time, the extent to which those items activate overlapping channels will constrain the amount of information that can be successfully processed. To examine this, we used brain/behavior correlations in which we directly compared behavioral performance on a perceptual task with the amount of overlap amongst the neural channels used to support the items from the behavioral task. In Chapter 1, we found that the amount of information that could be encoded on a change detection task was correlated with the amount of channel overlap within occipitotemporal cortex, but not early visual regions such as V1-V3. In Chapter 2, we extend this finding by showing that the amount of information that could reach visual awareness in a masking paradigm was also

predicted by overlap amongst occipitotemporal, as well as occipitoparietal channels, but once again not in V1-V3. Finally, in Chapter 3, we sought to identify which particular channels were the most behaviorally relevant and found that virtually any part of higher-level visual cortex (e.g. across occipitotemporal cortex, within category selective regions, within the least active voxels, amongst a random sample of voxels, etc.) was significantly correlated with behavioral performance. Together, these results suggest that visual cognition is limited by a set of neural channels that extend across the majority of higher-level visual cortex. These findings have direct implications on many prominent models of visual cognition, specifically those focused on perceptual limitations, and help clarify the large-scale representational structure in higher-level visual cortex.

Table of Contents

This thesis is dedicated to S.M.A., Y.I.K., & R.S.

Acknowledgements

Thanks to my advisors, Ken Nakayama and George Alvarez, who have given me more than they could possibly realize.

Thanks also to my longtime collaborator (and unofficial mentor), Talia Konkle, from whom I have learned so much.

Thanks to Daniel Dennett, Jeremy Wolfe, and Todd Horowitz for giving me my start in research. I can safely say I would not have ever made it this far without them.

Thanks to Sarah Cormiea, Morgan Henry, Julie Rhee, and Mirta Stantic for help with data collection and collaboration on many projects.

Thanks to the many people who gave me feedback about various aspects of this work: Nancy Kanwisher, Daniel Schacter, Arash Afraz, Bruno Breitmeyer, Patrick Cavanagh, Marvin Chun, Daniel Dennett, James DiCarlo, Sabine Kastner, Steve Luck, Amy Skerry, Rebecca Saxe, Nicholas Turk-Browne, Daniel Yamins, and Jeremy Wolfe. Special thanks to Maryam Vaziri Pashkam, who I can never fully thank for all of her time and generosity.

Thanks to the entire Vision Sciences Laboratory: Yaoda Xu, Sam Anthony, Eve Ayeroff, Julie Belkova, Katie Bettencourt, Tim Brady, Sasen Cain, Jonathan Cant, Sarah Cohan, Joe DeGutis, Judy Fan, Daryl Fougnie, Lucia Garrido, Laura Germine, Jonathan Gill, Fred Halper, Jason Haberman, Su Keun Jeong, Justin Junge, Andy Leber, Bria Long, Camille Morvan, Marnix Naber, Irene Pepperberg, Sonia Poltoratski, Anna Shafter-Skelton, Viola Stormer, Charles Stromeyer, Roger Strong, Arin Turek, Ruosi Wang, Jeremy Wilmer, Jiedong Zhang, and Xiaoyu Zhang.

Thank you to Celia Raia and other members of the psychology department staff for help in navigating this process.

This work was supported by an NSF-GRFP (Michael Cohen), an NRSA (F32EY022863 to Talia Konkle), an NIH NEI-RO1 (EY01362 to Ken Nakayama), and an NSF CAREER (BCS-0953730 to George Alvarez).

Thanks to my family, who have always been there for me throughout my entire life.

Thanks to Andrew and Rosie for listening to my maniacal ramblings.

And thanks to Yoana, whose support has made all of this possible.

0

Introduction

0.1 Bottlenecks in the world.

When you pour water out of a bottle, the rate at which it flows is limited by the width of the bottle's opening. If the opening is wide, the water will pour out rather fast. If the opening is small enough, the water will trickle out one drop at a time. The gradually narrowing passageway through which the water flows is referred to as the "bottleneck." While it may have initially referred to a specific part of a bottle, the term "bottleneck" now broadly refers to any particular structure or process that limits the capacity of a system.

Bottlenecks can be found throughout the world. One of the most common bottlenecks in human society occurs on highways. Virtually every driver has had the experience of sitting in a traffic jam caused by the closing of a few lanes (**Figure 0**). In this case, the reduction of the highway from multiple lanes down to one is the bottleneck on regular traffic flow. Bottlenecks can also be seen outside of human society. For example, a "population bottleneck" refers to the

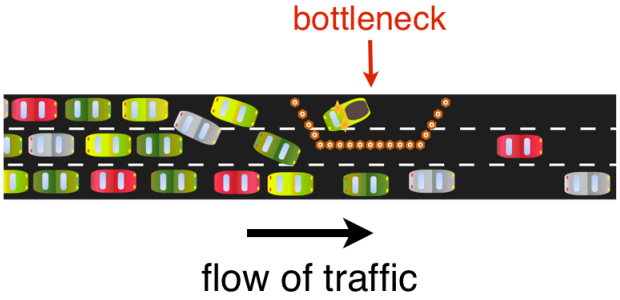

Figure 0: Example of a bottleneck in traffic.

sudden decrease in the size of a particular species due to a catastrophic event such as a drought or flood (Keller et al., 1994). The decrease in the population leads to a decrease in genetic diversity, which can weaken the fitness of that particular population. In this case, the reduction of the population from, say, several thousand to just a few dozen is the bottleneck on genetic evolution.

In psychology and neuroscience, it is often thought that different types of bottlenecks limit the processing capacity of the mind and brain. Such bottlenecks are necessary since the sensory environment contains more information than can be successfully processed (Chun & Wolfe, 2001). The most extensively studied of these bottlenecks is selective attention. Attention is the cognitive mechanism that selects some information for further processing at the expense of other, non-selected information (Broadbent, 1958). However, attention itself is a limited process, such that only a finite amount of information can be selected at a given moment. Thus, the limited capacity of attention is a bottleneck on human cognition.

0.2 High-level neural channels as a bottleneck on visual cognition

In this thesis, we identify another bottleneck on human cognition, specifically the processing capacity of the visual system. Rather than stemming from an attentional limitation, this particular bottleneck stems from the functional organization of the higher-level parts of the visual hierarchy. Over the past few decades, substantial progress has been made in characterizing the structures involved in representing different objects and categories (e.g. distributed activation patterns, category selective modules, etc.) (Haxby et al., 2001; Kanwisher, 2010). We suggest that these structures can be thought of as processing channels through which visual information is transmitted. Each channel is involved in representing/processing particular visual stimuli and

each has a limited representational capacity/bandwidth. Whenever multiple items have to be processed by the visual system, those items will often activate the same channels to varying degrees. Under this view, the extent to which different bits of information can be successfully processed is limited by the extent to which those bits activate the same channels. When there is relatively high overlap, less information will be processed since there will be competition within those processing channels. When there is less overlap, more information will be processed since different channels will be able to operate alongside one another.

To test this idea, we used brain/behavior correlations to directly compare perceptual capacity with the amount of overlap amongst neural channels. We behaviorally measured performance on a variety of tasks (e.g. change detection, visual masking, and visual search) with multiple visual categories and correlated these results with the extent to which those same categories activate overlapping neural channels. The overarching prediction is that there will be a significant relationship between performance on these behavioral tasks and channel overlap.

In order to determine what stimuli and neural regions to examine, we first re-considered what is already know about the functional organization of higher-level visual cortex.

The two pathways

Perhaps the most well established organizing principle of the visual system comes from the two-stream hypothesis. Under this view, after information is processed in primary visual cortex (V1), it is immediately transmitted into two basic processing streams: the ventral and dorsal streams (Ungerleider & Mishkin, 1982). The ventral stream is located along occipitotemporal cortex and is primarily comprised of inferotemporal (IT) cortex, while the dorsal stream takes up

most of the posterior part of the parietal cortex. In addition to being anatomically separated, these two pathways are also thought to be functionally distinct. Broadly speaking, the ventral stream is often considered to be the visual recognition pathway that binds together basic visual features (e.g. orientation, motion, color, etc.) to create distinct object/form representations. Meanwhile, the dorsal stream is considered to be the location/action pathway, which represents the spatial location of visual objects and helps mediate the coordination of visuo-motor actions (Goodale & Milner, 1992).

Ventral stream/occipitotemporal cortex

Within the ventral stream, decades of research have revealed several large-scale coding principles for representing different types of objects and categories. When considering the entire pathway, different categories of objects (e.g. faces, bodies, chairs, etc.) elicit unique activation patterns across large swath of cortex that can be decoded with sophisticated pattern-classification algorithms (Haxby et al., 2001; Tong & Pratte, 2012). One reason these algorithms are successful is that certain stimulus dimensions activate large-scale response patterns across a wide range of occipitotemporal cortex. Perhaps the most well known of these dimensions is the animate-inanimate distinction (Caramazza & Shelton, 1998; Martin, 2007). Animate categories (e.g. dogs, birds, deer, etc.) have been shown to primarily activate the lateral/ventral portions of occipitotemporal cortex, while inanimate categories (e.g. cars, chairs, beds, etc.) activate the more medial portions (Chao et al., 2002). Another dimension is real-world object size. In this case, big and small objects elicit distinct responses within the "inanimate region" (Konkle & Oliva, 2012).

Since this distinction is only seen within the inanimate region, animacy and real-world size together create a tripartite organization of occipitotemporal cortex (Konkle & Caramazza, 2013).

Within these large-scale networks, it has been repeatedly shown that the ventral pathway is comprised of several clusters of neurons that respond primarily to one particular category. Results from both human neuroimaging and monkey electrophysiology have identified clusters that respond to faces (Kanwisher et al., 1997; Tsao et al., 2006), bodies (Downing et al., 2001; Bell et al., 2009), scenes (Epstein & Kanwisher, 1998; Kornblith et al., 2013), written words (Dehaene & Cohen, 2011), and tools (Mahon et al., 2007). The clusters found at this smaller scale are often localized in a way that is consistent with the large-scale topography since the faces and body clusters are in the "animate region," while scenes are in the "big object" region (Konkle & Caramazza, 2013).

Dorsal stream/occipitoparietal cortex

While the ventral stream is classically thought to represent object form/identity, the dorsal pathway is thought to be primarily involved in representing objects' location and other information relevant for motor action. However, a growing body of evidence suggests that some object information is also encoded in the dorsal pathway. In single unit recordings in the macaque lateral intraparietal area, it has been repeatedly shown that certain neurons respond selectively to particular to two-dimensional geometric shapes (Sereno & Maunsel, 1998; Lehky & Sereno, 2007). This work has been recently extended into macaque anterior intraparietal area, in which shape selective neurons were identified after controlling for confounding factors such as object orientation and eye movements (Romero et al., 2014). Similar results have been found

using functional neuroimaging (fMRI) when using a passive viewing design (Chao & Martin, 2000; Sereno et al., 2002), fMRI adaptation (Konen & Kastner, 2008), or while having participants perform a visual working memory task (Xu & Chun, 2006). It is worth noting that while there are a growing number of studies reporting evidence consistent with the idea of object information being encoded in the dorsal pathway, these effects are almost always considerably weaker than those found in the ventral pathway.

0.3 Stimulus categories and basic neural measures

With these findings about the two visual processing streams in mind, we made several decisions about how we would test the idea that overlap amongst higher-level channels limits visual cognition. First, in terms of stimulus selection, we decided to use faces, bodies, scenes, and objects. We chose these stimuli because in addition to being represented by distinct category selective modules (Kanwisher, 2010), these categories also have reliable and distinct activation patterns outside of those modules (Haxby et al., 2001; Kriegeskorte et al., 2008). Thus, using these stimuli will allows us to compare and contrast how different coding structures relate to perceptual capacity. Second, when expanding the stimulus set beyond these four categories, we decided to include cars, cats, chairs, hammers, and phones. We chose these categories in order to sample from the dimensions that have been shown to drive large-scale cortical responses: large inanimate (cars and chairs), small inanimate (phones and hammers), animate (cats), and tools (hammers). Given that the neural responses elicited by animate/inanimate and big/small objects overlap with one another to varying degrees (Haxby, et al., 2001; Spiridon & Kanwisher, 2002; Kriegeskorte et al. 2008; Konkle & Caramazza, 2013), there will likely be enough variance in the

amount of overlap between categories for us to relate to behavior. In terms of our neural measure, we decided to obtain whole-brain responses to these categories so we could relate channel overlap across all parts of the visual hierarchy with behavioral performance. Specifically, a whole-brain measure enables us to compare brain/behavior correlations stemming from the ventral occipitotemporal cortex with those coming from the dorsal occipitoparietal cortex.

0.4 Plan of dissertation

In the first two Chapters of this dissertation, we examine the relationship between overlap amongst higher-level visual channels with performance on two types behavioral tasks. In Chapter 1, we showed that the extent to which items from four stimulus categories (faces, bodies, scenes, and objects) interfere with one another and limit performance on a change detection task is predicted by the overlap of the neural responses elicited by those same categories; more specifically, overlap within occipitotemporal cortex, but not occipitoparietal cortex, early visual cortex, or a prefrontal region associated with working memory (Curtis & D'Esposito, 2003). In Chapter 2, we continued to investigate the relationship between perception and neural overlap with a new task (visual masking) and a new set of stimuli and categories (bodies, buildings, cars, chairs, and faces). In this case, we found that the extent to which different stimulus categories mask one another (e.g. how well a face masks a building vs. how well a face masks a cat) was predicted by overlap amongst the neural patterns elicited by those categories. However, unlike in Chapter 1, we found correlations in both occipitotemporal and occipitoparietal cortex. The factors leading to a positive correlation in occipitoparietal cortex are discussed.

In Chapter 3, we turned our attention towards identifying the particular neural structures that are the most behaviorally relevant. To do this, we used a behavioral task (visual search) that yields quickly extremely reliable measures. Acquiring reliable data rather quickly is critical for this endeavor since it enabled us to increase the number of categories. We replicated ourselves once again in finding that overlap amongst higher-level neural channels predicts behavioral performance with strong correlations across both occipitotemporal and occipitoparietal cortex. However, we also made the surprising finding that the relationship between neural overlap and behavioral capacity could be found across virtually every part higher-level visual cortex. We found significant brain/behavior correlations when analyzing the large-scale sectors (e.g. occipitotemporal cortex.), category selective regions (FFA, PPA, EBA, LO), the most active voxels, the least active voxels, and even a random sample of voxels. Together, these results suggest that there is a ubiquitous and uniform representational structure across higher-level visual cortex.

The results from these Chapters require existent models of visual cognition to be modified, specifically models that aim to account for perceptual capacity limitations. In addition, the fact that the visual search data can be predicted across virtually any independent structure within higher-level visual cortex challenges certain accounts of what computations some of those structures are carrying out (e.g. FFA). Finally, these results open up a wide range of questions for future research: 1) Are there visual behaviors with high-level categories that cannot be predicted from our neural measure? 2) Does the relationship between neural structure and behavioral performance hold when analyzing individual trials/exemplars? Or does it rely on averaging across trials and examining the data at the category level? 3) Is this framework limited to the visual cognition, or is it a general bottleneck on cognition across all sensory modalities?

1

Processing multiple visual objects is limited by overlap in neural channels

1.0 Abstract

High-level visual categories (e.g., faces, bodies, scenes, and objects) have separable neural representations across the visual cortex. Here, we show that this division of neural resources affects the ability to simultaneously process multiple items. In a behavioral task, we found that performance was superior when items were drawn from different categories (e.g. two faces/two scenes) compared to when items were drawn from one category (e.g. four faces). The magnitude of this mixed-category benefit depended on which stimulus categories were paired together (e.g. faces and scenes showed a greater behavioral benefit than objects and scenes). Using functional neuroimaging (fMRI), we showed that the size of the mixed-category benefit was predicted by the amount of separation between neural response patterns, particularly within occipitotemporal cortex. These results suggest that the ability to process multiple items at once is limited by the extent to which those items are represented by separate neural populations.

1.1. Introduction

An influential idea in neuroscience is that there is an intrinsic relationship between cognitive capacity and neural organization. For example, seminal cognitive models claim there are distinct resources devoted to perceiving and remembering auditory and visual information (Baddeley & Hitch, 1974, Duncan, et al., 1997). This cognitive distinction is reflected in the separate cortical regions devoted to processing sensory information from each modality (Gazzaniga, et al., 2008). Similarly, within the domain of vision, when items are placed nearby each other, they interfere more than when they are spaced farther apart (Franconeri, et al., 2013; Pelli & Tillman, 2008). These behavioral effects have been linked to receptive fields and the retinotopic organization of early visual areas, in which items that are farther apart activate more separable neural populations (Wandel, et al., 2007; Desimone & Duncan, 1995; Kastner, et al., 2001). Thus, there are multiple cognitive domains in which it has been proposed that capacity limitations in behavior are intrinsically driven by competition for representation at the neural level (Franconeri, et al., 2013; Desimone & Duncan, 1995; Kastner, et al., 2001; Kastner, et al., 1998; Beck & Kastner, 2005).

However, in the realm of high-level vision, evidence linking neural organization to behavioral capacities is sparse, although neural findings suggest there may be opportunities for such a link. For example, results from fMRI and single-unit recording have found distinct clusters of neurons that selectively respond to categories such as faces, bodies, scenes, and objects (Kanwisher, 2010; Bell, et al., 2011). These categories also elicit distinctive activation patterns across the ventral stream as measured with fMRI (Haxby, et al., 2001; Kriegeskorte, et al., 2008). Together, these

results raise the interesting possibility that there are partially separate cognitive resources available for processing different object categories.

In contrast, many prominent theories of visual cognition do not consider the possibility that different categories are processed by different representational mechanisms. For example, most models of attention and working memory assume or imply that these processes are limited by content-independent mechanisms such as the number of items that can be represented (Pylyshyn & Storm, 1988; Drew & Vogel, 2008; Awh, et al., 2007; Zhang & Luck, 2008), the amount of information that can be processed (Alvarez & Cavanagh, 2004; Ma, et al., 2014; Franconeri, et al., 2007), or the degree of spatial interference between items (Franconeri, et al., 2013; Delvenne, 2005; Delvenne & Holt, 2012; Franconeri, et al., 2010). Similarly, classical accounts of object recognition are intended to apply equally to all object categories (Biederman, 1987; Tarr and Bülthoff, 1998). These approaches implicitly assume that visual cognition is limited by mechanisms that are not dependent on any major distinctions between objects.

Here, we examined (i) how high-level visual categories (faces, bodies, scenes, and objects) compete for representational resources in a change-detection task, and (ii) whether this competition is related to the separation of neural patterns across the cortex. To estimate the degree of competition between different categories, participants performed a task that required encoding multiple items at once from either the same category (e.g. four faces) or different categories (e.g. two faces and two scenes). Any benefit in behavioral performance for mixed-category conditions relative to same-category conditions would suggest that different object categories draw on partially separable representational resources. To relate these behavioral measures to neural organization, we used fMRI to measure the neural responses of these

categories individually and quantified the extent to which these categories activate different cortical regions.

Overall, we found evidence for separate representational resources for different object categories: performance with mixed-category displays was systematically better than performance with same-category displays. Critically, we also observed that the size of this mixed-category benefit was correlated with the degree to which items elicited distinct neural patterns, particularly within occipitotemporal cortex. These results support the view that a key limitation to simultaneously processing multiple high-level items is the extent to which those items are represented by non-overlapping neural channels within occipitotemporal cortex.

1.2. Results

1.2.1 A mixed-category benefit in behavior

To measure how items from different categories compete for representation, participants performed a task that required encoding multiple items at once. The stimulus set included images of faces, bodies, scenes, and objects matched for luminance and contrast (see **Figure A.1**). On each trial, four different items were presented for 800ms with one item in each visual quadrant. Following a blank display (1000ms), the items reappeared with one item cued by a red frame, and participants reported if that item had changed (**Figure 1a**). Changes occurred on half of the trials and could only occur at the cued location.

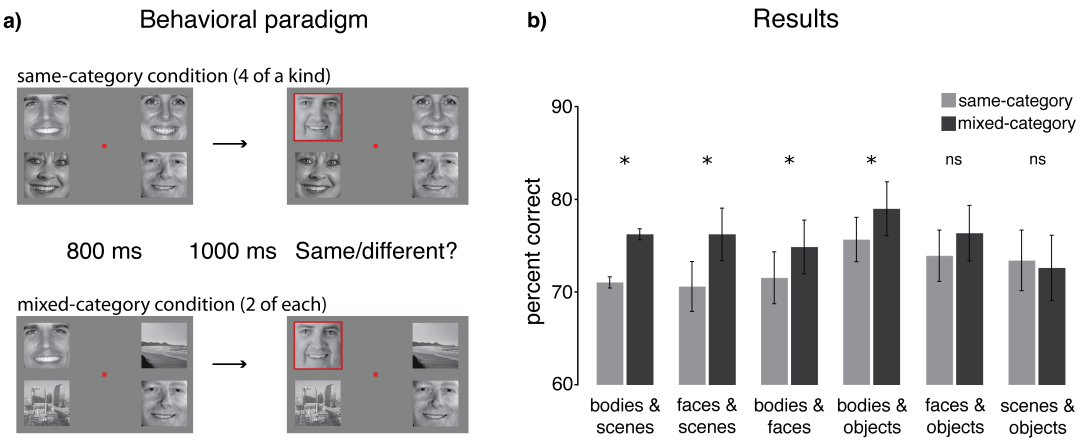

Figure 1: (**a**) Behavioral paradigm. Participants were shown two successively presented displays with four items in each display (**Methods**). On the second display, one item was cued (red frame) and participants reported if that item had changed. In the same-category condition, items came from the same stimulus category (e.g. four faces or four scenes). In the mixed-category conditions, items came from two different categories (e.g. two faces and two scenes). (**b**) Behavioral experiment results. Same-category (light grey) and mixed-category (dark grey) performance is plotted in terms of percent correct for all possible category pairings. Error bars reflect within-subject s.e.m (Loftus & Masson, 1994). *P<0.05.

The critical manipulation was that half of the trials were same-category displays (e.g. four faces or four scenes), while the other half of trials were mixed-category displays (e.g. two faces and two scenes). Whenever an item changed, it changed to another item from the same category (e.g. a face could change into another face, but not a scene). Each participant was assigned two categories, such as faces and scenes, or bodies and objects, in order to obtain within-subject measures for the same-category and mixed-category conditions. Across six groups of participants, all six pairwise combinations of category pairings were tested (**Methods**).

If items from one category are easier to process, participants might pay more attention to the easier category in mixed-category displays. To address this concern, we averaged performance across the tested categories (e.g., for the face-scene pair, we averaged over whether a face or scene was tested). Thus, any differences in overall performance for the mixed-category and same-category conditions cannot be explained by attentional bias towards one particular category. We

also took several steps to ensure that performance was approximately matched on the same-category conditions for all categories. First, we carefully selected our stimulus set based on a series of pilot studies (**Methods**). Second, before testing each participant, we used an adaptive calibration procedure to match performance on the same-category conditions, by adjusting the transparency of the items (**Methods**). Finally, we adopted a conservative exclusion criterion: any participants whose performance on the same-category displays (e.g. four faces compared to four scenes) differed by more than 10% were not included in the main analysis (**Methods**). This exclusion procedure ensured that there were no differences in difficulty between the same-category conditions for any pair of categories (P>0.16 for all pairings, **Figure A.2**). While these exclusion criteria were chosen to isolate competition between items, our overall behavioral pattern and its relationship to neural activation is similar with and without behavioral subjects excluded (**Figure A.3**).

Overall, we observed a mixed-category benefit: performance on mixed-category displays was superior to performance on same-category displays ($F_{1,9}$=19.85, P<0.01) (**Figure 1b**).This suggests that different categories draw on separate resources, improving processing of mixed-category displays. Moreover, while there was a general benefit for mixing any two categories, a closer examination suggested that the effect size depended on which categories were paired together (regression model comparison: $F_{5,54=}$2.29, P=0.059, **Methods**). The mixed-category benefit for each pairing, in order from biggest to smallest, was: Bodies-Scenes: 5.6%, SEM=1.5%; Faces-Scenes: 5.2%, SEM= 1.3%; Bodies-Faces: 3.3%, SEM=1.1%; Bodies-Objects: 3.3%, SEM=1.2%; Faces-Objects: 2.4%, SEM=1.9%; Scenes-Objects: -0.8%, SEM=1.9% (**Figure 1b**). The variation in the size of the mixed-category benefit suggests that categories do not compete

equally for representation and that there are graded benefits depending on the particular combination of categories.

What is the source of the variability in the size of the mixed-category benefit? We hypothesize that visual object information is represented by a set of broadly tuned neural channels in the visual system, and that each stimulus category activates a subset of those channels (Desimone & Duncan, 1995; Kastner, et al., 2001; Kastner, et al., 1998; Beck & Kastner, 2005; Wong, et al., 2008; Olsson & Poom, 2005). Under this view, items compete for representation to the extent that they activate overlapping channels. The differences in the size of the mixed-category benefit may thus be due to the extent to which the channels representing different categories are separated.

Importantly, while this representational-competition framework explains why varying degrees of mixed-category benefits occur, it cannot make a priori predictions about why particular categories (e.g. faces and scenes) interfere with one another less than other categories (e.g. objects and scenes). Thus, we sought to (i) directly measure the neural responses to each stimulus category and (ii) use these neural responses to predict the size of the mixed-category benefit between categories. Furthermore, by assuming a model of representational competition in the brain, we can leverage the graded pattern of behavioral mixed-category benefits to gain insight into the plausible sites of competition at the neural level.

1.2.2 Measuring neural separation among category responses

Six new participants who did not participate in the behavioral experiment were recruited for the fMRI experiment. Participants viewed stimuli presented in a blocked design, with each block

composed of images from a single category presented in a single quadrant (**Methods**). The same image set was used in both the behavioral and fMRI experiments. Neural response patterns were measured separately for each category in each quadrant of the visual field. There are two key properties of this fMRI design. First, any successful brain/behavior relationship requires that behavioral interference between two categories can be predicted from the neural responses to those categories measured in isolation and across separate locations. Second, by using two groups of participants, one for behavioral measurements and another for neural measurements, any brain/behavior relationship cannot rely on individual idiosyncrasies in object processing and instead reflects a more general property of object representation in behavior and neural coding.

To probe how different neural regions predict behavioral interference, we divided the set of visually active voxels into four sectors in each participant: occipitotemporal, occipitoparietal, early visual (V1-V3), and prefrontal cortex (PFC). These sectors were defined from independent functional localizers (**Methods** and **Figure A.4**). Performing the analysis in these sectors allowed us to examine the neural response patterns across the major divisions of the visual system: early retinotopic cortex, the what pathway, the where/how pathway (Ungerleider & Mishkin, 1982; Goodale & Milner, 1992), as well as in a frontal lobe region associated with working memory (Curtis & D'Esposito, 2003).

We defined the *neural separation* between any two categories as the extent to which the stimuli activate different voxels. To quantify this, we first identified the most active voxels in each sector for each of the categories, at a particular threshold (e.g. the 10% most active voxels for objects [objects>rest] and the top 10% most active voxels for scenes [scenes>rest], depicted in **Fig. 2**).

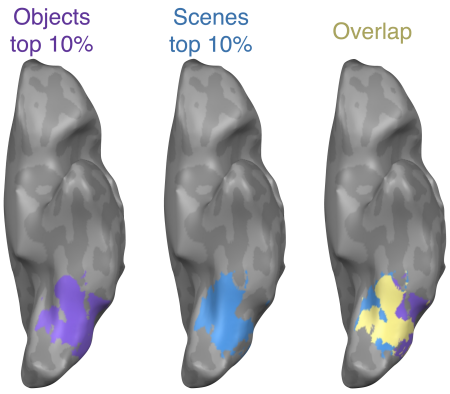

Figure 2: Visualization of the neural separation procedure. Activation and overlap among the 10% most active voxels for objects and scenes in the occipitotemporal sector is shown in a representative subject. The 10% most active voxels for each category are colored as: objects-purple; scenes-blue. The overlap among these active voxels are shown in yellow. For visualization purposes this figure shows the most active voxels and overlapping voxels combined across all locations; for the main analysis, overlap was computed separately for each pair of locations.

Next, we calculated the percentage of those voxels that were shared by each category pairing (i.e. percent overlap). This percent overlap measure was then converted to a neural separation measure as 1 - percent overlap. The amount of neural separation for every category pairing was then calculated at every activation threshold (from 1%-99%) (**Methods**). Varying the percent of active voxels under consideration allows us to probe whether a sparse or more distributed pool of representational channels best predicts the behavioral effect. This was done in every sector of each fMRI participant. In addition, we also used an area under the curve (AUC) analysis, which integrates over all possible activation thresholds, to compute an aggregate neural separation measure for each category pairing. Finally, we performed a standard representational similarity analysis in which the neural patterns of each category pairing were compared using a pattern dissimilarity measure (1 - the Pearson correlation between two response patterns across the entire sector (Kriegeskorte, et al., 2008; **Methods**).

1.2.3 Neural separation predicts the mixed-category benefit

To assess the degree to which neural separation predicted the mixed-category benefit, we correlated the amount of neural separation for every category pairing in each individual fMRI participant with the size of the mixed-category benefits from the behavioral experiments (i.e. a random effects analysis of the distribution of brain/behavior correlations; see **Methods**). An illustration of this analysis procedure using data obtained from occipitotemporal cortex is shown in **Figure 3.** We chose this analysis because it allows for a stronger inference about the generality of our results relative to a fixed effects analysis on both the neural and behavioral data (Kriegeskorte, et al., 2008). In addition, we were confident in our ability to analyze each fMRI participant individually given the highly reliable nature of our neural data (average within-subject split-half reliability in occipitotemporal: r=0.82, occipitoparietal, r=0.79, early visual: r=0.86, prefrontal: r=0.65; see **Figures A.5 & A.6**).

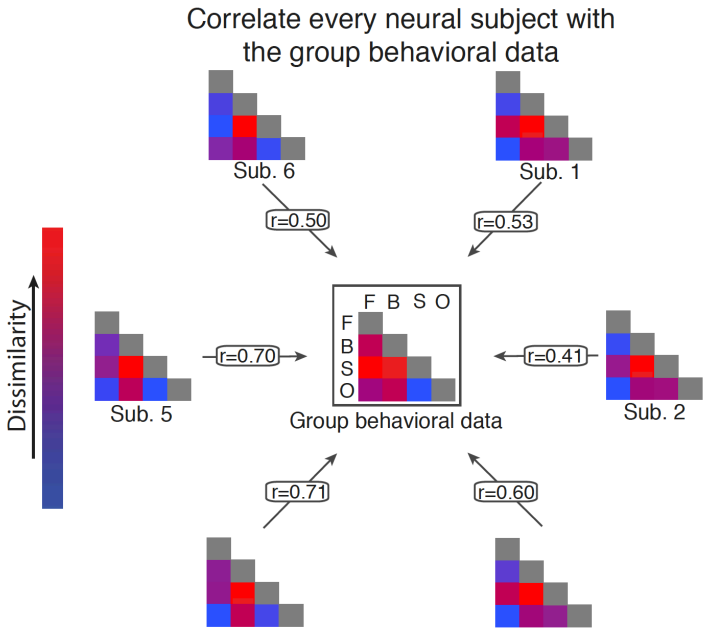

Figure 3: Visualization of our analysis procedure. The middle matrix represents the group data from the behavioral experiment with the color of each square corresponding to the size of the mixed-category benefit for that category pairing (see **Figure 1**). The six remaining matrices correspond to each fMRI participant with the color of each square corresponding to the amount of neural separation between two categories in occipitotemporal cortex at the 10% activation threshold (**Figure 2**). The correlations (*r*) are shown for each fMRI participant. Note that the *r*-values shown here are the same as those shown in **Figure 4b** for occipitotemporal cortex.

To assess whether the most active voxels alone could predict the mixed-category benefit, we correlated the amount of neural separation in the 10% most active voxels with the size of the mixed-category benefit. In this case, we found a significant correlation in occipitotemporal cortex of each participant (average r=0.59, P<0.01), with a smaller, but still significant correlation in occipitoparietal cortex (average r=0.30, P<0.01), and no correlation in early visual (average r=-0.03, P=0.82) or prefrontal cortex (average r=-0.06, P=0.63) (**Figure 4b).** A leave-one-category-out analysis confirmed that the correlations in each of these sectors were not driven by any particular category (**Methods**). It should be noted that, given the fine-grained retinotopy in early visual cortex, objects presented across visual quadrants activate nearly completely separate

regions, and this is reflected in the neural separation measure (ranging from 93% to 96% separation). Thus, by design, we anticipated that the neural separation of these patterns in the early visual cortex would not correlate with the behavioral results.

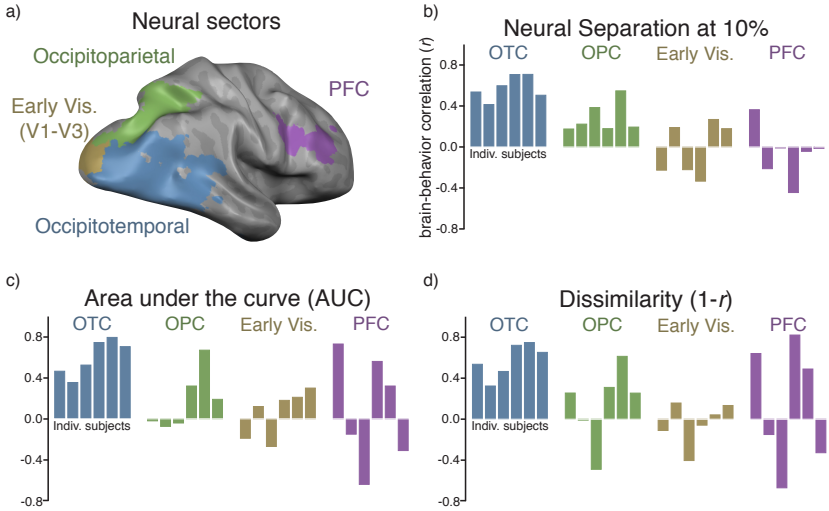

Figure 4: a) Visualization of the four sectors from a representative subject. b) Brain/behavior correlations in every sector for each fMRI participant at the 10% activation threshold, with *r*-values plotted on the y-axis. Each bar corresponds to an individual participant. c) Brain/behavior correlations in every sector for each participant when using the AUC analysis and d) representational dissimilarity (1 – *r*).

To compare correlations between any two sectors, a paired *t*-test was performed on the Fisher-*z* transformed correlations. In this case, the correlation in occipitotemporal cortex was significantly greater than the correlations in all other sectors (occipitoparietal: t_5=5.14, P<0.01, early visual: t_5=4.68, P<0.01; PFC: t_5=4.67, P<0.01). Together these results show that the degree of neural overlap between stimulus categories, particularly within occipitotemporal cortex, strongly predicts the variation in the size of the behavioral mixed-category benefit for different categories. Moreover, since this analysis only considers 10% of voxels in a given sector, these results indicate that a relatively sparse set of representational channels predicts the behavioral effect.

Next, we varied the activation threshold to test whether a more restricted or expansive pool

of neural channels could best predict the graded patterns of the behavioral mix-effect. The percent of most-active voxels used for the separation analysis was systematically increased from 1%–99% (at 100%, there is complete overlap between all pairs of categories since every voxel is included for every category). The brain/behavior correlation of every subject as a function of percentile for each sector is shown in **Figure 5**.

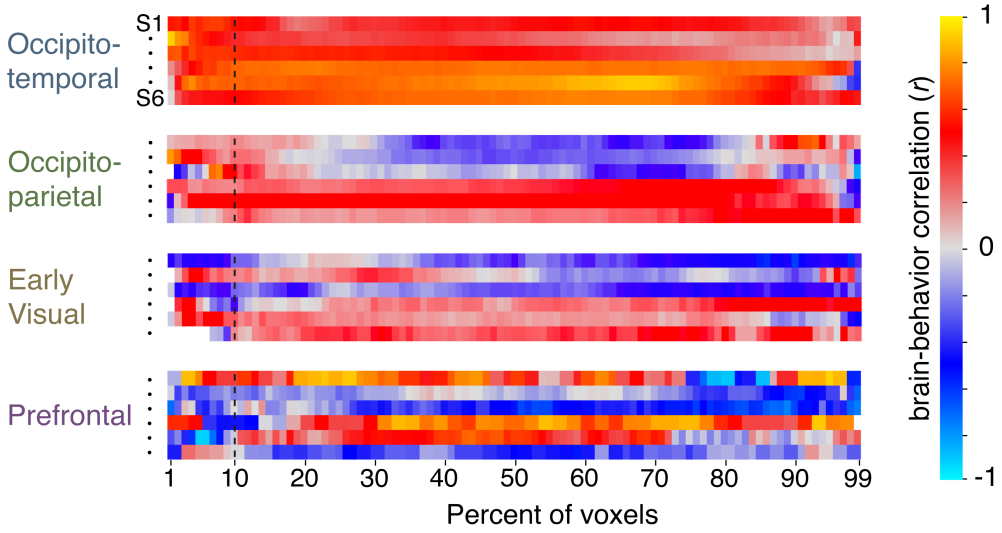

Figure 5: Brain/behavior correlation in each subject for every sector for all possible activation thresholds. Each row shows the results for one of the six individual subjects (in the same order as shown in **Figure 4** for each sector). The x-axis corresponds to the percentage of active voxels considered for the neural overlap analysis. The dashed vertical line marks the brain/behavior correlation when considering the 10% most active voxels, corresponding to the data in **Figure 4b**.

This analysis revealed that the behavioral effect is well-predicted by the amount of neural separation across a broad range of occipitotemporal cortex, regardless of the percentile chosen for the neural separation analysis. Thus, while separation within the most active voxels of occipitotemporal cortex can predict the mixed-category benefit, peaking with the 4% most active voxels, considering a larger pool of less activate voxels does not dramatically change the brain/behavior correlation until the bottom 30% of active voxels is reached (**Figure A.7**). Put

another way, for any two categories, the degree of overlap among the most active voxels is similarly reflected across the majority of the entire activation profile.

To assess the statistical reliability of this result in a way that does not depend on a particular activation threshold, we used an area under the curve (AUC) analysis to compute the aggregate neural separation between categories, for all subjects in all sectors. These values were then correlated with the behavioral results, and the results were similar to those observed when considering only the top 10% most active voxels (**Figure 4c**). There was a significant correlation in occipitotemporal cortex (average r=0.62, t_5=6.28, P<0.01), while little to no correlation was observed in the remaining sectors (occipitoparietal: r=0.20, early visual: r=0.06, prefrontal: r=0.11; P>0.21 in all cases) (**Figure 4c**). In addition, the correlation in occipitotemporal cortex was significantly greater than the correlations in all other sectors (occipitoparietal: t_5=8.36, P<0.001, early visual: t_5=6.86, P<0.01), except PFC where the effect was marginal (t_5=2.30, P=0.07), likely because the brain/behavior correlation measures in PFC were highly inconsistent across participants. It is worth noting that in the dorsal stream sector of occipitoparietal cortex, while there was a significant correlation at the 10% cutoff, the AUC analysis did not show a significant correlation (t_5=1.45, P>0.21), suggesting that only the most active voxels in occipitoparietal cortex predict the behavioral data.

A convergent pattern of results was found using a representational similarity analysis (14). That is, in occipitotemporal cortex, pattern dissimilarity (1 - r) across all pairs of categories strongly predicted the magnitude of the mixed-category benefit (average r=0.60, t_5=6.77, P<0.01; **Figure 4d**). None of the other sectors showed a significant brain/behavior correlation using this neural measure (P>0.37 in all cases), and direct comparisons between sectors show that the

brain/behavior correlation in occipitotemporal cortex was significantly greater than those in the other sectors ($P<0.05$ in all cases except in PFC, where the difference was not significant: $t_5=1.87$, $P=0.12$, again, likely due to the brain/behavior correlations being highly variable in PFC).

To what extent do the category-selective regions for faces, bodies, scenes, and objects found in the occipitotemporal sector drive these results (Kanwisher, 2010)? To address this question, we calculated the brain/behavior correlation in occipitotemporal cortex when considering only category selective regions (e.g. FFA/OFA and PPA/RSC, when comparing faces and scenes, etc.) or only the cortex outside the category selective regions, using pattern dissimilarity as our measure of representational similarity (**Methods**). This analysis revealed a strong brain/behavior correlation both within the category selective regions (average $r=0.62$, $P<0.01$) and outside the category selective regions ($r=0.60$, $P<0.01$) (**Figure A.8**), with no difference between these two correlations ($t_5=0.10$, $P=0.92$).

Different assumptions about neural coding are tested by our two analyses: neural separation tests the idea that information is conveyed by maximal neural responses; neural similarity assumes that information is conveyed over the full distribution of responses within some circumscribed cortical territory. These measures dissociate in the occipitoparietal sector. The neural overlap analysis revealed that only the most active voxels have systematic differences amongst categories, suggesting that there is reliable object category information along portions of this sector (Konen & Kastner, 2008). This observation was missed by the representational similarity analysis, presumably because many of the dorsal stream voxels are not as informative, making the full neural patterns subject to more noise. This result also highlights that the selection of voxels over which pattern analysis is performed can be critical to the outcomes. In contrast, in

the occipitotemporal cortex, both the separation and similarity metrics strongly correlated with behavior, and thus cannot distinguish between the functional roles of strong overall responses and. distributed patterns. Nevertheless, this convergence strongly demonstrates that neural responses across the entire occipitotemporal cortex have the requisite representational structure to predict object-processing capacity in behavior.

1.3 General Discussion

Here we characterized participants' ability to simultaneously process multiple high-level items and linked this behavioral capacity to the underlying neural representations of these items. Participants performed better in a change-detection task when items were from different categories than when items were from the same category. This suggests that within the domain of high-level vision, items do not compete equally for representation. Using fMRI to independently measure the pattern of activity evoked by each category, we found that the magnitude of the mixed-category benefit for any category pairing was strongly predicted by the amount of neural separation between those categories in occipitotemporal cortex. These data suggest that processing multiple items at once is limited by the extent to which those items draw on the same underlying representational resources.

The present behavioral results challenge many influential models of attention and working memory. These models are typically derived from studies that use simple stimuli (e.g. colorful dots or geometric shapes), and tend to posit general limits that are assumed to apply to all items equally. For example, some models propose that processing capacity is set by a fixed number of pointers (Pylyshyn & Storm, 1988; Drew & Vogel, 2008; Awh, et al., 2007; Zhang & Luck, 2008),

fixed resource limits (Alvarez & Cavanagh, 2004; Ma, et al., 2014; Franconeri, et al., 2007), or from spatial interference between items (Franconeri, et al., 2013; Delvenne, 2005; Delvenne & Holt, 2012; Franconeri, et al., 2010), none of which are assumed to depend on the particular items being processed. However, the present results demonstrate that the ability to process multiple items at once is greater when the items are from different categories. We interpret this finding in terms of partially separate representational resources available for processing different types of high-level stimuli. However, an alternative possibility is that these effects depend on processing overlap instead of representational overlap. For example, it has been argued that car experts show greater interference between cars and faces in a perceptual task than non-experts because only experts use holistic processing to recognize both cars and faces (McKeef, et al., 2010; McGugin, et al., 2011). The current behavioral data cannot distinguish between these possibilities. Future work will be required to determine which stages of perceptual processing show interference and whether this interference is best characterized in terms of representational or processing overlap.

Given that the size of the behavioral mixed-category benefit varied as a function of which categories were being processed, what is the source of this variability? We found that neural responses patterns, particularly in occipiotemporal cortex, strongly predicted the pattern of behavioral interference. This relationship between object processing and the structure of occipitotemporal cortex is intuitive since occipitotemporal cortex is known to respond to high-level object and shape information (Kanwisher, 2010; Bell, et al., 2011; Haxby, et al., 2001; Kriegeskorte, et al., 2008) and has receptive fields large enough to encompass multiple items in our experimental design (Kravitz, et al., 2008). However, some aspects of the correlation between

behavior and occipitotemporal cortex were somewhat surprising. In particular, we found that the relative separation between stimulus categories was consistent across the entire response profile along occipitotemporal cortex. That is, the brain/behavioral relationship held when considering the most active voxels, the most selective voxels (e.g. FFA/PPA), or those voxels outside of classical category selective regions.

The fact that the brain/behavior correlation can be seen across a large majority of occipitotemporal cortex is not predicted by expertise-based (Gauthier, et al., 2000) or modular (Kanwisher, 2010) models of object representation. If this correlation was due to differences in expertise between the categories, one might expect to see a significant correlation only in FFA (Tarr & Gauthier, 2000). Similarly, if competition within the most category-selective voxels drove the behavioral result, we would expect to only find a significant brain/behavior correlation within these regions. Of course, it is important to emphasize that the present approach is correlational, and so we do not know whether all or some subset of occipitotemporal cortex is the underlying site of interference. Future work using causal methods (Afraz, et al., 2006), or exploring individual differences in capacity and neural organization (Park, et al., 2010) will be important to explore these hypotheses.

In light of the relationship between behavioral performance and neural separation, it is important to consider the level of representation at which the competition occurs. For example, items might interfere with each other within a semantic (Moores, et al., 2003), categorical (Kanwisher, 2010), or perceptual (Ullman, 2006; Tanaka, 2003) space. Variants of the current task could be used to isolate the levels of representation involved in the mixed-category benefit and its relationship to neural responses. For example, using exemplars with significant perceptual

variation (e.g. caricatures, Mooney faces, and photographs) would better isolate a category level of representation. Conversely, examining the same type of brain/behavioral relationship within a single category would minimize the variation in semantic space and would target a more perceptual space. While the current data cannot isolate the level of representation at which competition occurs, it is possible that neural competition could occur at all levels of representation, and that behavioral performance will ultimately be limited by competition at the level of representation that is most relevant for the current task.

The idea of competition for representation is a prominent component of the biased competition model of visual attention, which was originally developed based on neurophysiological studies in monkeys (Desimone & Duncan, 1985), and has been expanded to explain certain human neuroimaging and behavioral results (Scalf, et al., 2013). These previous studies have shown that if two items are presented close enough to land within a single receptive field, they compete for neural representation, such that the neural response to both items matches a weighted average of the response to each individual item alone (Reynolds & Desimone, 1999). When attention is directed to one of the items, neural responses are biased towards the attended item, causing the neuron to fire as if only the attended item were present (Desimone & Duncan, 1995; Kastner, et al., 2001; Kastner, et al., 1998; Beck & Kastner, 2005; Reynolds & Desimone, 1999). In the current study, we did not measure neural competition directly. Instead, we measured neural responses to items presented in isolation and used similarity in those responses to predict performance in a behavioral task. We suggest that the cost for neural similarity reflects a form of competition, but we cannot say how that competition manifests itself (e.g. either as suppression of overall responses, or a disruption in the pattern of responses across

the population) or if these mechanisms are the same from early to high-level visual cortex (Treue, et al., 2000; Caradini, et al., 1998; MacEvoy, et al., 2009). Thus, parameterizing neural similarity and measuring neural responses to items presented simultaneously will be essential for addressing the relationship between neural similarity and neural competition.

Overall, the present findings support a framework in which visual processing relies on a large set of broadly tuned coding channels, and perceptual interference between items depends on the degree to which they activate overlapping channels. This proposal predicts that a behavioral mixed-category benefit will be obtained for tasks that require processing multiple items at once, to the extent that the items rely on separate channels. It is widely known that there are severe high-level constraints on our ability to attend to, keep track of, and remember information. The present work adds a structural constraint on information processing and perceptual representation, based on how high-level object information is represented within the visual system.

1.4 Methods

Behavioral Task.

Participants (N=55) viewed 4 items for 800ms, followed by a fixation screen for 1000ms, followed by a probe display of 4 items, one of which was cued with a red frame. For any display, images were randomly chosen from the stimulus set with the constraint that all images on a given display were unique. There were no changes on half the trials. On the other trials, the cued item changed to a different item from the same category (e.g., from one face to another). Participants

reported if the cued item had changed. The probe display remained on screen until participants gave a response using the keyboard.

On same-category trials, all four items came from the same category (e.g., four faces, or four scenes), with each category appearing equally often across trials. On mixed-category trials, there were two items from each category (e.g., two faces and two scenes). Items were arranged such that one item from each category appeared in each visual hemifield. Across the mixed-category trials, both categories were tested equally often. The location of the cued item was chosen in a pseudo-random fashion.

Stimuli

For all behavioral experiments, stimuli were presented on a 24-inch LCD monitor with a 60 Hz refresh rate. Participants sat approximately 57 cm away from the monitor such that 1° of visual angle subtended 1 cm on the display. Experiments were created and controlled on a computer running MATLAB with the Psychophysics Toolbox (1, 2). Images were presented at 6° x 6° of visual angle, with a different image appearing in each quadrant of the visual field, 8.4° away from fixation. Within a hemifield, the center-to-center distance of items was approximately 7.5°, while the center-to-center distance of two items that were in different hemifields but on the same horizontal plane was 15.4°. A red fixation dot (.55°) was presented in the middle of the display. The background of the display had an average luminance of 73.8 cd/m^2.

Stimulus selection.

Stimuli were chosen in an attempt to minimize the possibility that participants could perform the task by focusing on low- or mid-level features. All items we matched for luminance and contrast. All faces were vertically oriented, looking at the camera, smiling, Caucasian, approximately between the ages of 20-50, and each image was cropped so the outer contours of the head/hair were not showing. Thus, participants would have to represent the "face" rather than notice a change in outer contour or a difference in hair texture. Body images were of a single individual (author MAC) wearing the same outfit in a variety of action poses (i.e. jumping, throwing, etc.). Scenes were a variety of natural scene categories that never contained faces, bodies, or any of the objects from the object category. Objects were specifically selected to have a similar, round outer contour. This was done in an attempt to force participants to focus on the object itself (e.g. an apple or a clock), rather than just focus on the bounding contour alone.

Procedure

Each participant initially performed a calibration block of same-category trials with their assigned categories (e.g. four faces, four scenes). In this block, there were 30 practice trials and 120 experimental trials. Half of the experimental trials were allocated for each of the two assigned categories. The QUEST staircase (Watson & Pelli, 1983) procedure was used to adaptively change the transparency of the items until performance on same-category displays for both categories was approximately 70% (actual same-category performance across all six conditions was 73.83%). These transparency thresholds were then used for the stimuli in the main experiment. In the main experiment, participants performed 20 practice trials followed by four blocks of 80

experimental trials (320 total experimental trials). Within a block, all trial types (i.e. same-category/mixed-category, different types of mixed-category display configurations, etc.) were shuffled and appeared in a random order.

Participants & exclusion criteria

In order to obtain N=10 for each of the 6 category pairings (e.g. faces-scenes, faces-bodies, etc.), 55 participants completed one or more of the category-pairings experiments. The visibility matching procedure described above was designed to ensure that performance was equal on both the same-category conditions. However, in some cases, individuals showed large differences in performance across the two categories that could not be eliminated with this procedure. Thus, participants whose performance on the two same-category conditions differed by more than 10% were excluded. To ensure that this procedure did not result in an atypical sample, we correlated the size of the mixed-category benefit when the exclusion criterion was and was not applied. There was a very strong correlation between these data sets (r=0.93) with the different mixed-category benefits as follows: Faces & Scenes: with exclusion = 5.2%, without exclusion = 5.1%; Bodies and Scenes with exclusion = 5.6%, without exclusion = 5.2%; Bodies and Faces with exclusion = 3.3%, without exclusion = 3.2%; Bodies and Objects with exclusion = 3.3%, without exclusion = 4.2%; Faces and Objects with exclusion = 2.4%, without exclusion = 2.6%; Objects and Scenes with exclusion = -0.8%, without exclusion = 2.1%). See **Fig S5** for the different brain/behavior correlations in occipitotemporal cortex using the top 10% and AUC measurements with versus without excluded participants.

Mixed-Category Benefit Analysis

The graded nature of the mixed category benefit was evaluated by comparing regression models where category-pair was included as a factor (i.e., a separate parameter was fit for each category pair) versus not included as a factor (i.e., a null model with an intercept term only). Model comparison was conducted using an ANOVA, to test whether the model that estimated the mix effect separately for each category pair fit the data reliably better than a model that fit a single intercept across all category pairs.

fMRI Task.

6 participants (none of whom performed the behavioral task) completed the fMRI experiment. There were 8 runs of the main experiment, 1 run of meridian mapping to localize early visual areas (V1-V3), 1 run of a working memory task used to localize PFC, and 2 localizer runs used for defining the occipitotemporal and occipitoparietal sectors as well as category-selective regions (FFA/OFA, PPA/RSC, EBA/FBA, and LO).

Experimental Runs: Participants were shown the same faces, bodies, scenes, and objects that were used in the behavioral experiments (**Figure A.1**), presented one at a time in each of the 4 quadrants (top-left, top-right, bottom-left, bottom-right), following a standard blocked design. Locations are subsequently abbreviated as TL (top left), TR (top right), BL (bottom left), BR (bottom right). In each 16s stimulus block, images from one category were presented in isolation (one at a time) at one of the 4 locations; 10s fixation blocks intervened between each stimulus block. 11 items were presented per block for 1s with a 0.45 s intervening blank. Participants were instructed to maintain fixation on a central cross and to press a button indicating when the same

item appeared twice in a row, which happened once per block. For any given run, all 4 stimulus categories were presented in two of the four possible locations, for two separate blocks per category x location pair, yielding 16 blocks per run. Across 8 runs, the categories were presented at each pair of locations (TL-TR; TL-BR; BL-BR; BL-TR), yielding 8 blocks for each of the 16 category x location conditions (see below for information on localizer runs).

Neural separation analysis

The logic of this analysis is to compute the proportion of voxels that are activated by any two categories (e.g., faces and scenes): if no voxels are co-activated, then there is 100% neural separation, whereas if all voxels are co-activated, then there is 0% separation. This analysis relies on one free parameter, which sets the percent of the most active voxels to consider as the available representational resources of each object category. In addition, we take into account location by considering the overlap between the two categories at all pairs of locations, and then averaging across location pairs.

To compute the neural separation between two categories within a sector (e.g. faces and scenes in occipitotemporal cortex), we used the following procedure: (1) The response (betas) for each category-location pair were sorted and the top N% were selected for analysis, where the N was varied from 1% to 99%. (2) Percent overlap at a particular threshold was computed as the number of voxels that were shared between any two conditions at that threshold, divided by the number of selected voxels for each condition (e.g. if 10 voxels overlap among the top 100 face voxels and the top 100 scene voxels, the face-scene overlap would be 10/100 = 10%). To take into account location, percent overlap was calculated separately for all 12 possible location pairs:

FacesTL-ScenesTR, FacesTL-ScenesBL, FacesTL-ScenesBR, etc. **Figure 2** shows the example where N=10% for the activation patterns of objects (purple), scenes (blue), and shared (yellow). (3) Finally, we averaged across these 12 overlap estimates to compute the final overall estimate of overlap between a pair of categories. This measure can be interpreted as the degree to which two different categories in two different locations will recruit similar cortical territory. We computed percent overlap at each percentile (i.e. 1% through 99% of the most active voxels), generating a neural overlap curve, and converted this to a %Separation measure by taking 1-%Overlap. This procedure was conducted for all pairs of categories, for all sectors, for all subjects.

fMRI Acquisition

Structural and functional imaging data were collected on a 3T Siemens Trio scanner at the Harvard University Center for Brain Sciences. Structural data were obtained in 176 axial slices with 1 x 1 x 1 mm voxel resolution, TR = 2200 ms. Functional blood oxygenation level-dependent (BOLD) data were obtained using a gradient-echo echo-planar pulse sequence (33 axial slices parallel to the anterior commissure-posterior commissure line; 70 x 70 matrix; FoV = 256 x 256 mm; 3.1 x 3.1 x 3.1 mm voxel resolution; Gap thickness = 0.62 mm; TR = 2000 ms; TE = 60 ms; flip angle = 90 degrees). A 32-channel phased-array head coil was used. Stimuli were generated using the Psychophysics toolbox for MATLAB and displayed with an LCD projector onto a screen in the scanner that subjects viewed via a mirror attached to the head coil.

fMRI Experiment Localizer Runs

Meridian map runs: Participants were instructed to maintain fixation and were shown blocks of flickering black and white checkerboard wedge stimuli, oriented along either the vertical or horizontal meridian. The apex of each wedge was at fixation and the base extended to 8 degrees of visual angle in the periphery, with a width of 4.42 degrees. The checkerboard pattern flickered at 8 Hz. The run consisted of 4 vertical meridian and 4 horizontal meridian blocks. Each stimulus block was 12 s with a 12 s intervening blank period. The orientation of the stimuli (vertical vs. horizontal) alternated from one block to the other.

DLPFC Runs: Participants performed a change detection task in which they had to say if an item changed between displays. Each display consisted of four items with each item placed in one of the four visual quadrants. The configuration of items on these displays was not identical to the configuration used in the behavioral experiment (see **Figure 1**). In this case, all items were equidistant from fixation and each item was equally close to the two adjacent items within and across the visual hemifields. On every trial, all items came from the same category (e.g. four faces, four bodies, etc.), with the images and locations selected randomly. The first display appeared for 1 s, followed by a 0.7 s blank interval with a fixation cross, and then a second display was presented for 1 s. Participants responded during a 1.8 s inter-trial-interval. Immediately before each trial, the black fixation dot turned red to alert participants that the next trial was about to begin. A change detection block consisted of 8 trials, in which each of the four categories was used on 2 trials. Changes occurred on half of the trials such that a change occurred once with every category in every block. Each run was composed of 3 change detection blocks of 32 s each, with fixation periods of 32 s following each block.

Localizer Runs: Participants performed a one-back repetition detection task with blocks of faces, bodies, scenes, objects, and scrambled objects. Stimuli in these runs were different from those in the experimental runs. Each run consisted of 10 stimulus blocks of 16 s, with intervening 12 s blank periods. Each category presented twice per run, with the order of the stimulus blocks counterbalanced in a mirror reverse manner (e.g. face, body, scene, object, scrambled, scrambled, objects, scene, body, face). Within a block, each item was presented for 1 s followed by a 0.33 s blank. Additionally, these localizer runs contained an orthogonal motion manipulation: In half of the blocks, the items were presented statically at fixation. In the remaining half of the blocks, items moved from the center of the screen towards either one of the four quadrants or along the horizontal and vertical meridians at 2.05 degrees per second. Each category was presented in a moving and stationary block.

fMRI Data Analysis

All fMRI data was processed using Brain Voyager QX software (Brain Innovation, Mastricht, Netherlands). Preprocessing steps included 3D motion correction, slice scan-time correction, linear trend removal, temporal high-pass filtering (0.01 Hz cutoff), spatial smoothing (4mm FWHM Kernel), and transformation into Talairach space. Statistical analyses were based on the general linear model. All GLM analyses included box-car regressors for each stimulus block convolved with a gamma-function to approximate the idealized hemodynamic response. Motion correction regressors were included as regressors of no interest. For each experimental protocol, separate GLMs were computed for each participant and run, yielding beta maps for each condition.

Defining neural sectors

Sectors were defined in each participant using the following procedure. Using the localizer runs, a set of visually-active voxels was defined based on the contrast of [Faces + Bodies + Scenes + Objects] vs. Rest (FDR<0.05, cluster threshold 150 contiguous 1x1x1 voxels) within a gray matter mask. To divide these visually-responsive voxels into sectors, the early visual sector included all active voxels within V1, V2, and V3, which were defined by hand on an inflated surface representation based on the horizontal vs. vertical contrasts of the meridian mapping experiment. The occipitotemporal and occipitoparietal sector were then defined as all remaining active voxels (outside of early visua), where the division between the dorsal and ventral streams was drawn by hand in each participant starting at the edge of early visual and determined by the spatial profile of active voxels along the surface. Finally, the PFC runs were used to identify the PFC sector from the contrast of working memory vs. rest; this region was masked by the gray matter but not by visually active voxels (as for some subjects, no frontal voxels were significantly visually active).

Neural Dissimilarity Analysis

This analysis follows the representational similarity methods outlined in Krigeskorte et al., (2008). For each pair of categories we computed the Pearson correlation between the response patterns (betas) across the entire sector (e.g. the OTC), and converted this to a dissimilarity measure (1-r). Location was taken into account in the same way as described in the channel separation analysis (see Methods and Materials in the main text).

Category-Selective ROI Analysis

To define category-selective regions, we computed standard contrasts for face selectivity (faces >[bodies scenes objects]), scene selectivity (scenes >[bodies faces objects]), and body selectivity (bodies >[faces scenes objects]) based on independent localizer runs. For object-selective areas, the contrast of objects>scrambled was used. In each participant, face-, body-, scene- and object-selective regions were defined using a semi-automated procedure that selects all significant voxels within a 9 mm radius spherical ROI around the weighted center of category-selective clusters (Peelen & Downing, 2005), where the cluster is selected based on proximity to the typical anatomical location of each region based on a meta analysis. All ROIs for all participants were verified by eye and adjusted if necessary. Category-selective regions included FFA and FBA (faces), PPA and RSC (scenes), EBA and FBA (bodies), and LOC (objects).

We next computed neural dissimilarity inside the category-selective ROIs for all pairs of categories, considering only the ROIs specific for the two categories in each pair. For example, the correlation between the response patterns to scenes and the response patterns to faces was computed considering only the voxels within scene- and face-selective regions, and converted to a dissimilarity measures (1-*r*). This dissimilarity measure for faces and scenes was then compared to the size of the mixed-category benefit with faces and scenes. Location was taken into account following the same procedure described in the channel separation analysis. To compute neural dissimilarity outside the category-selective ROIs, the same procedure was followed, but the voxels under consideration for scene and face competition, for example, were all voxels in the OTC that were *not* in any of the scene- or face-selective regions.

Within-Subject Pattern Reliability

We computed the reliability of neural response patterns separately for each fMRI participant and each brain sector using the following procedure. The data was split into two halves, such that each half contained one run of every possible location-location pairing (i.e. each half had a [top left]-[bottom right] run, a [bottom left]-[bottom right] run, etc.). Next, the pattern for each category in each location was correlated between the two halves of the data (e.g., activity for faces in the top left for half 1 was correlated with activity for faces in the top left for half 2). This correlation was calculated for each category in each location. Averaged across categories, locations, and subjects, the overall correlations were quite high in each sector (**Figure A.5**), indicating that these neural patterns were highly reliable within subjects.

Between-Subject Consistency in Neural Overlap

In every sector of each participant, the amount of overlap (top 10% and AUC) or dissimilarity (1-*r*) was calculated for each category pairing. Those measurements across all category pairings (6 total) were then compared between all subjects. The correlation between each subject in every sector using all three similarity measures was quite high overall, indicating a high degree of between-subject consistency in neural overlap between categories (**Figure A.6)**.

Assessing the statistical significance of the brain/behavior correlations

All brain/behavior correlations were computed for each individual fMRI subject against the group behavioral data. For each fMRI subject, an *r*-value (Pearson's correlation value) was calculated using a particular neural similarity measure (i.e. neural separation, pattern

dissimilarity). The *r*-values obtained for a given brain/behavior correlation were Fisher-z transformed and then tested for statistical significance using a random effects analysis. To compare the relative strength of two given correlations (e.g. is the brain/behavior correlation stronger in occipitotemporal cortex compared to early visual areas), a within-subject's *t*-test was performed on the transformed correlation coefficients obtained from two regions.

It should be noted that in every subject, a brain/behavior correlation was obtained using the same group average behavioral data. However, despite the common behavioral measures, the brain/behavior correlation values across brain participants are still independent from one another and are not intrinsically correlated. Simulations were run to ensure the independence of such correlation values and more generally to validate the statistical procedure used. In these simulations, the behavioral data were fixed and the brain data were randomly set for each subject, the brain/behavior correlation was obtained for each subject, and the distribution of these brain/behavior correlations was assessed. The results of this simulation showed a null-distribution of correlation values centered around 0 that was well-approximated by a normal distribution following the Fisher-*z* transformation.

2

Overlap amongst visual processing channels limits visual awareness: Evidence from brain/behavior correlations

2.0 Abstract

Chapter 1 demonstrated that the ability to process multiple visual items is limited by the extent to which they activate overlapping neural channels. In this Chapter, we sought to extend this finding into another realm of visual cognition, specifically the limitations of visual awareness. We suggest that a source of these limitations is the organization of higher-level visual cortex into broadly tuned processing channels. Under this view, the extent to which items activate the same channels constrains the amount of information that can be processed by the visual system and ultimately reach awareness. To examine this, we measured how effectively items from different categories masked one another (e.g. cars masking bodies). In addition, we used fMRI to measure the similarity of neural responses elicited by these categories across the visual hierarchy. We found strong correlations between masking efficacy and neural similarity in higher-level regions across the ventral and dorsal pathways, but not in early visual areas (V1-V3). These results suggest that the organization of higher-level visual regions imposes an anatomical constraint on visual awareness and the overall processing capacity of visual cognition.

2.1 Introduction

Decades of research have revealed that the capacity of human visual awareness is surprisingly limited. Paradigms such as change blindness, inattentional blindness, the attentional blink, and visual crowding have demonstrated the numerous ways in which visual information fails to be consciously perceived. These limitations have often been attributed to the finite supply and spatial resolution of visual attention (Chun & Wolfe, 2001). Indeed, a variety of prominent models view attention as perhaps the primary factor that allows information to be accessed by awareness. Such models include the Global Neuronal Workspace (Dehaene & Changeux, 2011), the Local Recurrency Theory (Block, 2005; Lamme, 2010), Central Executive Theory (Broadbent, 1958), the Intermediate Level Theory (Jackendoff, 1987; Prinz, 2005), the Multiple Drafts Model (Dennett, 2005), and the Sensorimotor Theory (O'Regan & Noë, 2001). While attention undoubtedly plays a role in limiting visual awareness (Koch & Tsuchiya, 2007; Cohen & Dennett, 2011; Cohen, et al., 2012), there are other possible sources that have yet to be thoroughly examined.

Here, we suggest that visual awareness is limited by competition for representation, particularly in high-level processing channels. Under this view, stimuli that activate the same high-level channels will interfere with one another and limit what can be accessed by awareness. We examine the extent to which stimuli activate the same processing channels by using fMRI to measure the similarity of the neural response patterns elicited by different object categories across the visual hierarchy. It is well established that there are large-scale organizing principles across high-level visual cortex involved in representing visual categories such as faces, bodies, and objects (Kanwisher, 2010; Kravitz, et al., 2013), as well as broad superordinate categories

such as animacy and real-world size (Konkle & Oliva, 2012; Konkle & Caramazza, 2013). We suggest that these distinct neural structures shape the underlying cognitive architecture of the visual system and form multiple channels through which visual information is processed. Each channel has a finite processing capacity, and different stimuli elicit responses in these channels to varying degrees. The extent to which items activate the same channels constrains the processing capacity of the visual system and the amount of information ultimately available for conscious processing. When there is relatively high overlap in the activated channels, different bits of information will interfere in a mutually suppressive fashion and less information will be available for conscious processing. When there is less overlap, these channels can operate alongside one another with minimal interference, increasing the amount of information that can be accessed by consciousness. If this idea is correct, the degree to which information can reach awareness should correlate with the similarity of the neural responses involved in representing that information.

To examine this possibility, we performed two visual masking experiments in which items from five different categories — bodies, buildings, cars, chairs, and faces — served as both the target and the mask in all possible pairings of categories (Breitmeyer, 2007; Kouider & Dehaene, 2007). We used a Bayesian adaptive staircase procedure to estimate the presentation durations necessary for each category pairing (e.g. detecting a body among buildings, a car among faces, etc.) to result in equal behavioral performance across all conditions (80% accuracy). Overall, we found significant differences in how well different categories masked one another, with certain category pairings requiring more than twice as much time as others.

We then used fMRI to measure the similarity of the neural responses elicited by each category. With these neural responses, we asked if neural similarity predicts the behavioral results from the

masking experiment. Across both behavioral experiments, we found strong brain/behavior correlations in ventral occipitotemporal, lateral occipitotemporal, and occipitoparietal cortex, with the strongest correlations coming from ventral occipitotemporal cortex. Meanwhile, the correlations in early visual cortex (V1-V3) were either not significant (Experiment 1) or were significantly weaker than those in both occipitotemporal regions (Experiment 2).

These results suggest that the organization of the visual cortex across both the ventral and dorsal pathways limits the amount of information that can reach visual awareness. While a variety of models of awareness have prominent roles for certain cognitive mechanisms such as visual attention, our findings suggest that they must also account for the anatomical constraints described here. We suggest that attentional and anatomical limitations may come together under a unified framework in which overlap in higher-level visual cortex constrains the amount of information available not just for awareness, but also for attentional selection. Under this view, overlap amongst these neural channels is a processing bottleneck that precedes attention, with attention operating upon the information that has made its way through these channels. Such a framework could help link cognitive theories that posit the existence of distinct pools of attentional resources (Awh, et al., 2004; Alvarez, et al., 2004; Wickens, 2008) with the organization of the visual system. In this case, these pools of cognitive resources may be realized in the brain by the amount of separation amongst higher-level processing channels.

2.2 Experiment 1

2.2.1 Methods & Analyses

Participants

For the behavioral experiment, twenty participants gave informed consent and received either compensation or class credit for their participation. For the neuroimaging experiment, six participants gave informed consent and received compensation.

Stimuli

Stimuli were images of bodies, buildings, cars, chairs, and faces with 30 exemplars in each category. All items were selected to be as visually variable as possible. For example, the face set was comprised of people who were different ages, races, and genders, with variations in hairstyles, and looking directions with respect to the camera. All images were matched to have the same average luminance, contrast, and overall spectral energy (**Figure 6a** and **Methods**). These steps were taken to minimize the possibility of participants performing the task by simply relying on low-level features that might differentiate one stimulus category from another.

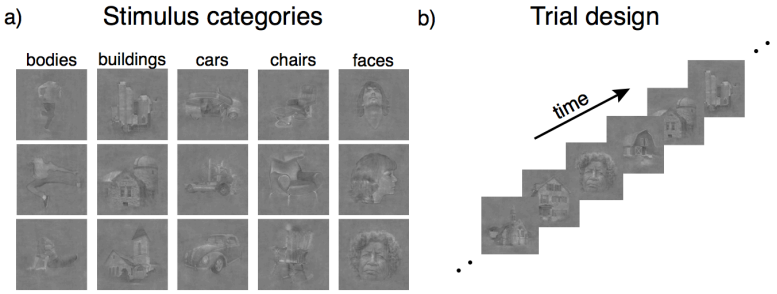

Figure 6: a) Examples of images from each of the five categories used in the behavioral and neuroimaging experiments. b) Representation of a sample trial in which one target — a face — is shown amongst nine distracting masks — buildings — presented in rapid succession at the center of the display. Presentation duration was adaptively set to achieve 80% correct detection performance.

Behavioral masking task

Participants performed a visual masking task in which they had to detect one target item amongst several rapidly presented forward- and backward-masking items. On every trial, ten total images were shown in immediate succession in the center of the display (**Figure 6b** and **Methods**). When a target was present, it was either the 4^{th}, 5^{th}, 6^{th}, or 7^{th} item presented on the display, with the target appearing in each position equally often. There were 10 experimental blocks with 10 practice trials and 96 experimental trials. Each block was defined by its particular category pairing (e.g. buildings and faces, cars and chairs, etc.) such that all 10 possible category pairings were measured in each subject. The order in which the category pairings were shown was counterbalanced with a Latin square design (Williams, 1949). Within a block, both categories served as the target and the masks equally often (e.g. buildings masking faces and faces masking buildings). A target appeared on half of the trials for both target/mask configurations. All trial types were randomly ordered within a block, so participants were unaware of what the target/mask configuration would be or if a target would be present. Thus, the task can be thought of as an "oddball" detection task.

We used QUEST, a Bayesian adaptive staircase procedure, to estimate the presentation duration for each target/mask configuration that would yield 80% behavioral performance (Watson & Pelli, 1983) (**Methods**). For every block, we interleaved separate staircases for each target/mask configuration, enabling us to obtain separate estimates of the threshold duration for each target/mask configuration. These estimates were averaged together to estimate the presentation duration for that particular category pairing (see **Appendix B** for analyses when the estimates are not averaged together).

Neuroimaging

Six participants who did not perform the masking task were scanned using fMRI. Within the experimental runs, participants passively viewed the same items used in the behavioral experiment in a blocked design, with each block composed of images from a single category (**Methods**). Participants' only task was to perform a simple vigilance task to press a button when a red circle appeared around an item. There were 4 runs of the main experiment, 1 run of a meridian map used to localize early visual cortex (V1-V3), and 2 localizer runs to define other neural sectors encompasing object-responsive cortex.

In each hemisphere of each participant, four neural sectors were defined to reflect the major divisions of the visual system. Early visual cortex was defined first based on the meridian mapping, with occipitotemporal (ventral stream) and occipitoparietal (dorsal stream) cortex defined as all visually active voxels beyond early visual cortex. Occipitotemporal cortex was then divided into ventral and lateral portions based on the occipitotemporal sulcus (see **Methods** for further information on all runs and procedures used to localize the four sectors).

Brain/behavior correlation analyses

To determine if masking efficacy is predicted by neural similarity, we performed a representational similarity analysis in which the neural response patterns of all categories were correlated with one another using the Pearson correlation (*r*) in each sector (Kriegeskorte, et al., 2008b). This analysis yields a full neural similarity matrix for all category pairings. We then asked which sector's neural similarity structure best predicted the behavioral similarity structure as

measured by the presentation duration needed to obtain equal behavioral performance for each category pairing.

To assess the statistical significance of the brain/behavior correlations, we carried out two types of analyses: First, we performed a permutation analysis on the group-level data, which reflects fixed effects of both the behavioral and neural measures (Kriegeskorte, et al., 2008b; **Methods**). Second, we used a linear mixed effects (LME) modeling, a more sensitive measure that estimates the brain/behavior correlations with random effects of both the behavioral and neural participants (Barr, et al., 2013; Winter, 2013; **Methods**). Both of these analyses were used to assess the overall significance of the brain/behavior correlations. However, only the LME method was used to compare the strength of the correlations between sectors since this method allows us to test for a within-subject effect of our brain regions (e.g. ventral occipitotemporal vs. early visual cortex) while simultaneously generalizing across both behavioral and neuroimaging participants.

2.2.2 Results

In the masking experiment, there were significant differences in the estimated presentation durations needed to equate behavioral performance across the different category pairings ($F_{1,9}$=15.93, P<0.001, η_p^2 =0.93) (**Figure 7**). An extreme example of this difference is that cars and chairs (95 ms/item) had to be presented almost twice as long as buildings and faces (48 ms/item) for equal performance.

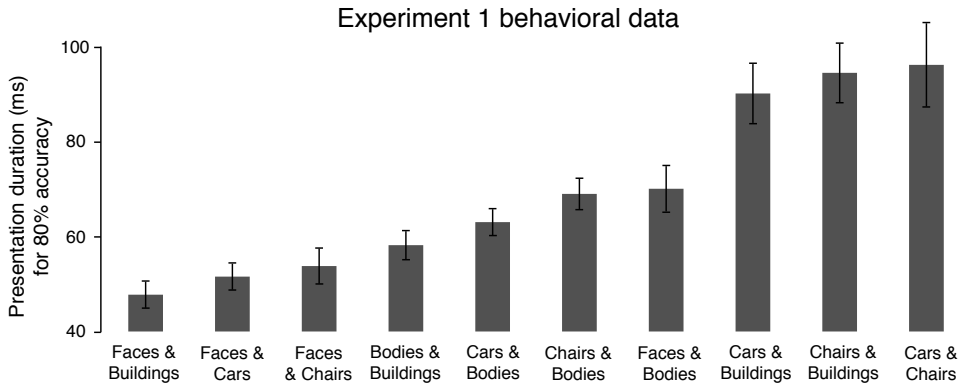

Figure 7: Results from the behavioral experiment. Estimated presentation durations needed to result in equal performance for each category pairing are shown on the y-axis. Each bar corresponds to a particular category pairing. Error bars reflect within-subject s.e.m.

What accounts for the variation in the masking efficacy between different category pairings? One possibility is that in spite of our efforts to eliminate low-level differences between the categories, enough differences remained such that some category pairings had more similar low-level features than others. This would be consistent with numerous results showing that an effective mask will be similar to the target in terms of low-level features (Hellige, et al., 1979; Legge & Foley, 1980; Michaels & Turvey, 1979). Alternatively, it is possible that the differences between the category pairings reflect differing degrees of overlap in higher-level neural/processing channels.

To distinguish between these possibilities, we correlated the behavioral results with the neural similarity values measured in each of the four sectors. If the differences in masking efficacy were due to low-level similarity, we would expect to find the strongest brain/behavior correlations in early visual areas (V1-V3), which contain neurons that respond to features such as orientation, spatial frequency, and simple feature combinations (Freeman, et al., 2013). Alternatively, if overlap amongst higher-level feature combinations determines these masking differences, we

should see the strongest correlations in the occipitotemporal cortex, and potentially occipitoparietal cortex, where it has been suggested that complex visual shape information is also encoded (Konen & Kastner, 2008).

Using a group-level permutation analysis, we found strong correlations in ventral occipitotemporal (r=0.84, P<0.001), lateral occipitotemporal (r=0.69, P<0.05), and occipitoparietal cortex (r=0.52, P<0.05), consistent with the notion that the differences in the masking task were due to similarity in neural response patterns (**Figure 8b**). Meanwhile, there was no correlation in early visual cortex (r=0.05, P=0.44), suggesting that lower-level similarity between these images did not drive the masking differences.

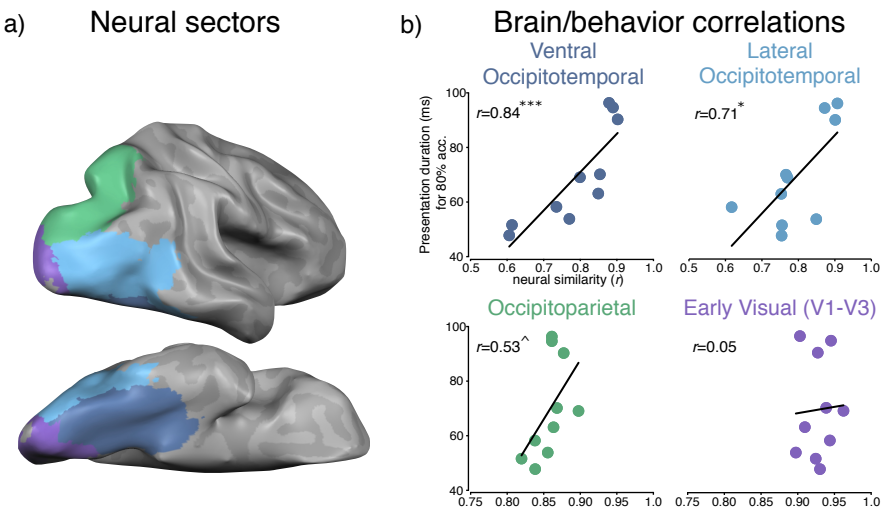

Figure 8: a) Visualization of the four neural sectors from a representative participant. b) Group level brain/behavior correlations in all sectors in Experiment 1. Neural similarity between category pairings plotted on the x-axis. Estimated presentation durations for equal behavioral performance for all category pairings plotted on the y-axis. Each dot corresponds to a single category pairing (e.g. faces and buildings). The values on both axes were calculated by averaging across all behavioral and neuroimaging participants respectively and plotted accordingly. Note the change in the scales of the x-axis between the two occipitotemporal and occopitoparietal/early visual plots. ***P<0.001, *P<0.05, ^P<0.10.

A convergent pattern of results was obtained with the linear mixed effects (LME) analysis. For this analysis, the masking results from every individual behavioral participant (N=20) were

correlated with the neural similarity values obtained in every neuroimaging participant (N=6). The results of every brain/behavior correlation for all participant-by-participant pairings (N=120) are shown in **Figure 9**. It is worth noting that in ventral occipitotemporal cortex, every combination of behavioral and neuroimaging participants (120 out of 120; **09**, top row) yielded brain/behavior correlations greater than zero, and in lateral occipitotemporal all but six participant combinations (114 out of 120; **Figure 9**, second row) resulted in correlations greater than zero. We modeled the Fisher *z*-transformed correlation values as a function of the neural sector, including random effects of behavioral and neuroimaging participants on both the intercept and slope term of the model (Barr, et al., 2013) (**Methods**).

The results of this analysis found significant brain/behavior correlations in ventral occipitotemporal (LME parameter estimate=0.74, t=9.76, P<0.001) lateral occipitotemporal (parameter estimate=0.49, t=6.44, P<0.001), and occipitoparietal cortex (parameter estimate=0.32, t=2.54, P<0.05), but not in early visual cortex (parameter estimate=0.08, t=0.83, P=0.40). In addition, we also used the LME analysis to compare the strength of the correlations in the different sectors. To do this, we measured the differences in the slope term of the linear mixed effects model for two given sectors to see if the slope in one sector was significantly different than that in another. This revealed that the correlations in ventral occipitotemporal cortex were greater than those in the three other sectors (slope estimates <-0.25, t<-2.30, P<0.05 in all cases), the correlations in lateral occipitotemporal cortex were significantly different from those in early visual cortex (slope estimates=-.41, t=-3.00, P<0.01), and the correlations in occipitoparietal were trending, but not significantly different from early visual cortex (slope

estimates=-.24, t=-1.65, P=0.09). Meanwhile, the correlations in lateral occipitotemporal and occipitoparietal cortex were not significantly different (slope estimates=-0.16, t<-1.41, P=0.16).

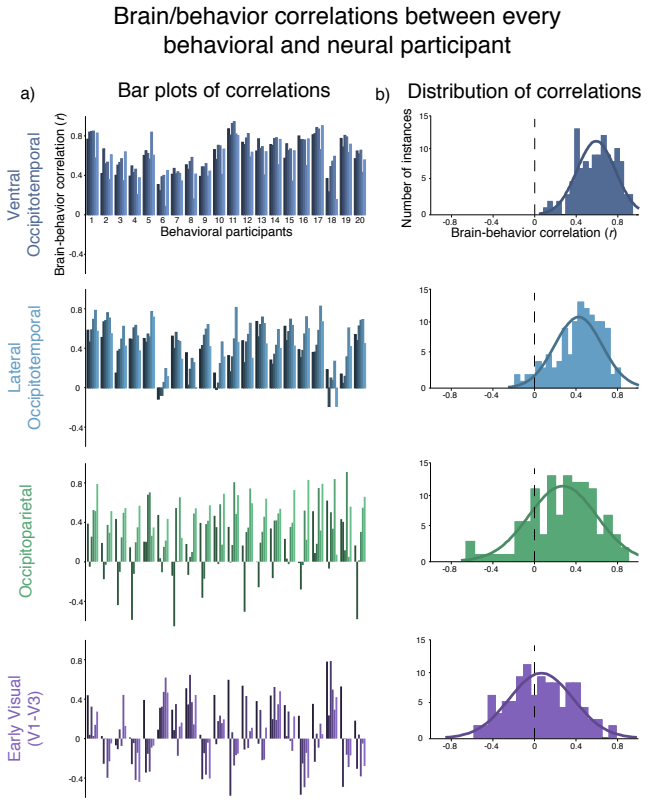

Figure 9: Brain/behavior correlations in all sectors for every possible behavior-neuroimaging participant combination from Experiment 1. a) Brain/behavior correlations are plotted on the y-axis. Each individual behavioral participant is plotted as unique clusters containing six bars on the x-axis. Within each cluster, the six differently shaded color bars represent that behavioral subject's brain/behavior correlations with the six neuroimaging participants. b) The same data in the bar graphs plotted as histograms fitted with a normal density function with the brain/behavior correlations on the y-axis. The number of instances of a particular brain/behavior correlation are plotted on the x-axis. Average correlations and standard deviations of the four sectors: ventral occipitotemporal: r=0.60, st. dev.=0.08; lateral occipitotempora: r=0.42, st. dev.=0.08; occipitoparietal: r=0.28, st. dev.=0.13; early visual: 0.07, st. dev.=0.09.

2.2.3 Discussion

Together, these findings support the notion that the amount of information that can be accessed by awareness is limited by the organization of the visual cortex. For the stimuli and categories tested here, it appears that the ventral and lateral occipitotemporal cortex place particularly strong constraints on awareness. While there may be instances in which overlap amongst low-level features and early visual areas constrain visual processing (Hellige, et al., 1979; Legge & Foley, 1980; Michaels & Turvey, 1979), we found no evidence of such a lower-level bottleneck: the correlations between early visual areas and behavior were not significant and were significantly less than those correlation from ventral and lateral occipitotemporal cortex. Instead, these results are consistent with the idea that processing these stimuli was limited by competition within higher-level visual pathways, which limits the amount of information that can reach visual awareness.

2.3 Experiment 2

2.3.1 Methods & Analyses

In Experiment 1, participants did not know what the target/mask configuration would be from one trial to the next within a block (e.g. in a face-building block, the target could be a face amongst building masks or vice versa). We chose this design so that participants could not tune their attention to one particular category and suppress the other, forcing the two categories to interfere with one another in a more equitable fashion (Desimone and Duncan, 1995). In Experiment 2, we tested whether the overall pattern of results would change if participants knew

what the target and mask categories were ahead of time (e.g. "In this session, faces will be the targets, and buildings will be the masks").

Participants

Twenty new behavioral participants who did not participate in Experiment 1 gave informed consent and received compensation or class credit for their participation. The six previous neuroimaging participants and their data were used again for Experiment 2.

Behavioral masking task

Experiment 2 followed the same procedure as Experiment 1 but differed in the following respects. Participants were told in advance to detect a particular target category in the item sequence. Each of the 10 category-pairing blocks was divided into two sections for each target category (e.g. face targets among building masks in the first section, building targets among face masks in the second). Participants were informed at the beginning of each section of the particular target/masking configuration. For each block, the order of the two configuration sections was randomly determined across participants. The practice trials at the beginning of each block were also presented in this order. At the mid-way point, instructions appeared on the screen alerting participants to the change in the target/mask configuration (**Methods**).

Neuroimaging & brain/behavior correlation analyses

The neuroimaging data from Experiment 1 was used in Experiment 2, and all analyses procedures were the same.

2.3.2 Results

For the masking task, the measurements from both configurations were averaged together for each category pairing in the same way as was done in Experiment 1 (see **Appendix B** for analysis when the estimates were not averaged together). Consistent with the previous results, there were once again significant differences between the category pairings in the presentation durations estimated to result in equal performance ($F_{1,9}$=8.20, P<0.001, η_p^2 =0.87). In fact, the behavioral results between the two experiments were highly correlated (r=0.88, P<0.001, **Appendix B**).

For the group-level data, a permutation analysis revealed a pattern of brain/behavior correlations similar to those in Experiment 1. The correlation in ventral occipitotemporal was very strong (r=0.69, P<0.05), with marginally significant correlations in both lateral occipitotemporal, (r=0.51, P=0.06) and occipitoparietal cortex (r=0.46, P<0.09) (**Figure 10b**). The correlation in early visual cortex was not significant (r=0.31, P=0.18).

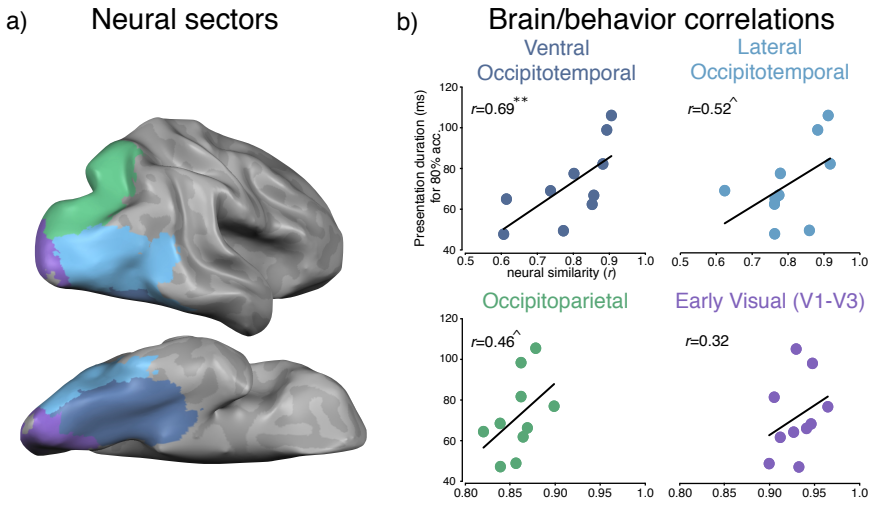

Figure 10: a) Visualization of the four neural sectors from a representative participant (same as show in Figure 3). b) Group level brain/behavior correlations in all sectors in Experiment 2. Neural similarity between category pairings plotted on the x-axis. Estimated presentation durations plotted on the y-axis. Each dot corresponds to a single category pairing (e.g. faces and buildings). Note the change in the scales of the x-axis between the two occipitotemporal and occopitoparietal/early visual plots. **P<0.01, ^P<0.10

The linear mixed effect model converged on the same pattern of results with one interesting exception (**Figure 11**). As before, we found significant correlations in ventral occipitotemporal (parameter estimate=0.60, t=8.71, P<0.001), lateral occipitotemporal (parameter estimate=0.42, t=5.57, P<0.001), and occipitoparietal cortex (parameter estimate=0.30, t=2.34, P<0.05). However, unlike what was seen with the permutation analysis, we also found a significant brain/behavior correlation in early visual cortex (parameter estimate=0.20, t=2.69, P<0.01). This difference is likely due to the fact that the LME analysis is a more sensitive measure that takes into account each individual behavioral and neural participant, while the permutation analysis averages across participants and only considers fixed-effects within the group-level data.

Comparing the strength of the correlations between the sectors, we once again found that the correlations within ventral occipitotemporal cortex were greater than all other sectors (slope estimates <-0.18, t<-2.05, P<0.05 in all cases). In addition, the correlation within lateral occipitotemporal cortex was marginally greater than early visual cortex (slope estimate=-0.22, t=-1.84, P=0.07). Meanwhile, there was no difference in the correlations between lateral occipitotemporal and occipitoparietal cortex (slope estimate=-0.12, t=-1.17, P=0.25) or occipitoparietal and early visual cortex (slope estimate=-0.10, t=-0.76, P=0.45).

Finally, we tested if the strength of the brain/behavior correlations different in ventral occipitotemporal and early visual cortex between the two behavioral experiments. In the case of ventral occipitotemporal cortex, the correlation observed in Experiment 1 was marginally greater than that observed in Experiment 2 (parameter estimate=-0.15, t=-1.95, P=0.051). Similarly, the difference in the correlations in early visual cortex for the two experiments was also trending, but not significant (parameter estimate=-0.12, t=-1.60, P=0.10). Thus, there does not appear to be a

significant trade-off between ventral occipitotemporal and early visual cortex between the two experiments. Though it is possible that such a trade-off could be detected with more statistical power.

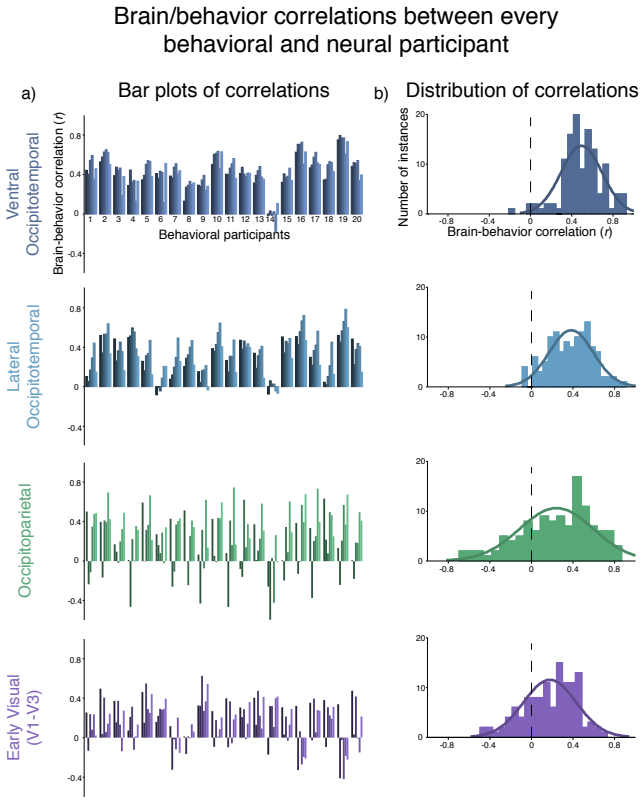

Figure 11: Brain/behavior correlations in all sectors for every possible behavior/neuroimaging participant combination from Experiment 2. Brain/behavior correlations are plotted on the y-axis. Each individual behavioral participant is plotted as unique clusters containing six bars on the x-axis. Within each cluster, the six differently shaded color bars represent that behavioral subject's brain/behavior correlations with the six neuroimaging participants. Average correlations and standard deviations of the four sectors: ventral occipitotemporal cortex: *r*=0.50, st. dev.=0.07; lateral occipitotemporal: *r*=0.37, st. dev.=0.08; occipitotemporal: *r*=0.26, st. dev.=0.13; early visual: *r*=0.18, st. dev.=0.07.

2.3.3 Discussion

Consistent with Experiment 1, the best predictor of the masking results was neural similarity within ventral and lateral occipitotemporal cortex. However, unlike Experiment 1, there was also a significant brain/behavior correlation within early visual cortex, though it was not significantly greater than the correlation in the previous experiment. One possible explanation for this result is that when participants are aware of the particular target category they are looking for, they are able to tune their attention to certain lower-level features that aid in detecting the target. For example, when buildings are the targets and faces are the masks, participants can tune their attention to vertical lines that are not regularly present in faces but are extremely common in buildings. It should be noted that even though such an attentional strategy may explain the emergence of a significant correlation in early visual cortex, that correlation is still less than those found within ventral and lateral occipitotemporal cortex. Thus, even when participants may tune their attention to focus on particular lower-level features, similarity within a higher-level feature space seems to play a larger role in determining the amount of information eventually available for conscious processing.

2.4 General Discussion

Here, we report several brain/behavior correlations between the masking efficacy of different category pairings and their neural similarity in higher-level visual cortex. Meanwhile, no significant correlations were found in early visual cortex (Experiment 1) unless participants were aware of the target category and could potentially focus on specific lower-level features that

would help differentiate one category of stimuli from another (Experiment 2). However, even when the correlation in early visual cortex was significant, it remained smaller than those found in ventral and lateral occipitotemporal cortex. Thus, we suggest that a particularly potent limitation of visual awareness stems from the large-scale organization of ventral aspect of the occipitotemporal cortex, with progressively weaker contributions of the lateral occipitotemporal, occipitoparietal, and early visual cortex, respectively. Previous work discussing what limits access to awareness has overwhelmingly focused on spatial or featural attention (Broadbent, 1958; Jackendoff, 1987; O'Regan & Noë, 2001; Dennett, 2005; Block, 2005; Prinz, 2005; Lamme, 2010; Dehaene & Changeux, 2011). While we do not dispute the critical role that attention may play (Koch & Tsuchiya, 2007; Cohen & Dennett, 2011; Cohen, et al., 2012), we suggest that there are architectural constraints (neural pathways) that are not accounted for in these models.

How does the organization of these neural channels limit perception during the masking task? Since images were not masked during the fMRI experiment and were consciously perceived by participants, it is natural to wonder about the mechanics of the behavioral result in terms of these neural channels. We suggest that visual information can only be accessed by consciousness if there is widespread recurrent/reentrant processing between the higher-level neural channels involved in encoding the information (i.e. ventral and lateral occipitotemporal cortex) and the prefronto-parietal networks (Lamme, 2010; Dehaene & Changeux, 2011). The presentation of a mask prevents recurrent/reentrant processing (Di Lollo, et al., 2000; Del Cul, et al., 2007), even if the early part of the neuronal response is largely unaffected (Kovacs, et al., 1995). When there is relatively high overlap amongst the channels activated by two stimulus categories (e.g. cars & chairs), the mask will be more effective at preventing further downstream processing since it will

be processed by the same set of channels involved in the recurrent/reentrant processing of the target. When the presentation time between these items increases, there is more time for this recurrent/reentrant processing before the mask interferes. The fact that the neural responses elicited by items that are consciously perceived predicts how well those categories will mask one another highlights the ways in which mapping the large scale organization of these higher-level neural/processing channels can be useful in understanding what factors constrain conscious processing.

If attention and these neural pathways are two distinct bottlenecks on conscious access, it is natural to wonder how they might interact with one another in a single framework. We suggest that processing channels in higher-level visual cortex may constrain the amount of information available for both visual awareness, as well as attentional selection. Under this view, attention selects information only after it has made its way through these neural/processing channels. Thus, overlap amongst these channels is an information-processing limit that precedes visual attention. If this is correct, this framework could help clarify a variety of behavioral results used to support the notion of multiple resource pools involved in cognitive processing (Awh, et al., 2004; Alvarez, et al., 2004; Wickens, 2008). In this case, the extent to which there are different resource pools is directly linked to the extent to which there is separation amongst these neural channels. Indeed, one of the strengths of this framework is that it could aid in understanding how the notion of multiple resources is realized in the brain by illuminating the extent to which different attentional/resource pools relate to the functional and anatomical organization of the cerebral cortex (Franconeri, et al., 2013).

If this notion about the link between neural structure and attentional resources is correct, it would alter the interpretation of previous results that have focused exclusively on attention. For example, it has been claimed that perceiving animals or vehicles requires little or no attention (Li, et al., 2002, but see Cohen, et al., 2011). This claim is based on the fact that detecting an animal/vehicle is unaffected by simultaneously searching for a T amongst Ls. However, under the framework described here, one reason there is no interference between the tasks could be that they are likely represented by significantly different neural channels (Peelen & Kastner, 2011; Dehaene & Cohen, 2011). If the search task also involved images of animals/vehicles, there may be a cost to adding another task since both tasks would rely on the same neural mechanisms. In that case, claims about the relationship between attention and awareness would need to be clarified since such findings might reflect the anatomical constraints described here, rather than the scope and limits of attention.

Going forward, future work will need to fully characterize the computations involved in each of these neural sectors to properly understand the constraints they impose on awareness. For example, one unexpected finding here is that neural similarity within occipitoparietal cortex, a region not generally thought to contain object information, predicts masking efficacy (Kravitz, et al., 2013). One possibility is that these correlations indicate the existence of object information within the dorsal pathway (Konen & Kasnter, 2008). However, it has recently been shown that damage to the ventral pathway reduces object-specific activity within the dorsal pathway (Konen, et al., 2011), which suggests that object representations in the dorsal stream are dependent upon the ventral stream. In that case, the correlations within occipitoparietal cortex may be related to

the transfer of information from the ventral stream into the parieto-prefrontal network that is often closely associated with conscious access (Dehaene & Changeux, 2011).

Overall, these results suggest that the limits of visual awareness are partly due to the organization of the higher-level visual system. It is possible that this bottleneck is not limited to visual awareness, and that sensory awareness in all modalities is limited by the organization of the relevant neural pathways. Anatomical constraints of conscious awareness may be a ubiquitous phenomenon, which limits conscious processing in all domains of cognition.

2.5 Methods

For both behavioral experiments, stimuli were presented on a 15.5-inch Nanao FlexScan T2-17ts monitor with a 120 Hz refresh rate and a screen resolution of 800x600. Participants sat approximately 57 cm away from the display so that 1 cm on the display would correspond to 1° of visual angle. The experiments were created and controlled with MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Images used in both the behavioral and fMRI experiment were grayscaled and normalized on multiple low-level dimensions across the entirety of each image using the SHINE toolbox (Willenbockel, et al., 2010).

Experiment 1 Behavioral Methods

Ten images were presented successively in the middle of the display with no blank gaps between items. All stimuli were square 5.9° x 5.9° images. On each trial, a red fixation dot (~0.1°) would appear in the center of the screen for 500ms. Immediately afterwards, the fixation dot

would turn black and the ten images would be shown with the fixation dot remaining on every image. At the end of a trial, participants reported whether or not a target was present by pressing a button on the keyboard. Visual feedback was immediately given for 500ms. Participants had to press a key to proceed to the next trial.

During the experiment, a staircase procedure was used to adaptively change the presentation duration of all items based on the accuracy of participants' response. The initial presentation rate of the practice trials was 112 ms/item, while the initial presentation rate of the experimental trials was 64ms/item. Two independent staircases were interleaved within a block for each target/mask combination (e.g. face target/building mask vs. building target/face mask) and resulted in two estimates of the presentation rate that would yield 80% performance. The data from each block (i.e. the presentation rates and accuracy on each trial) were subsequently analyzed using QUEST to determine the final presentation duration estimates for both conditions within a block. Those estimates were then averaged together to form a "category pairing estimate" (see below for analyses when those estimates were not averaged together).

It should be noted that the refresh rate of the monitor (120 Hz) constrained the possible presentation rates. At 120 Hz, each frame is shown for 8ms, thus, as the staircase procedure changed the presentation duration in response to the participants' performance, changes were made in increments of 8ms.

Experiment 2 Behavioral Methods

The only difference between Experiments 1 and 2 is that participants were told within each block what the particular target/mask configuration would be. Participants were informed of the configuration before the practice trials, reminded of that configuration when the experimental trials began and then were alerted to the change in configuration halfway through the experimental trials (i.e. after trial 48). There were no practice trials for the second configuration within a block.

Neuroimaging experiment

fMRI Acquisition: Structural and functional imaging data were collected on a 3T Siemens Trio scanner at the Harvard University Center for Brain Sciences. Structural data were obtained in 176 axial slices with 1 x 1 x 1 mm voxel resolution, TR = 2200 ms. Functional blood oxygenation level-dependent (BOLD) data were obtained using a gradient-echo echo-planar pulse sequence (33 axial slices parallel to the anterior commissure-posterior commissure line; 70 x 70 matrix; FoV = 256 x 256 mm; 3.1 x 3.1 x 3.1 mm voxel resolution; Gap thickness = 0.62 mm; TR = 2000 ms; TE = 60 ms; flip angle = 90 degrees). A 32-channel phased-array head coil was used. Stimuli were generated using the Psychophysics toolbox for MATLAB and displayed with an LCD projector onto a screen in the scanner that subjects viewed via a mirror attached to the head coil.

6 participants (none of whom performed the behavioral masking task) completed the fMRI experiment. There were 4 runs of the main experiment, 1 run of meridian mapping to localize early visual areas (V1-V3), and 2 localizer runs used for defining the occipitotemporal and occipitoparietal sectors.

Experimental Runs –These runs were used as part of a broader project within our laboratory and thus more categories were presented during the fMRI experiment than were ultimately used for this project. Participants viewed images of nine categories: bodies, buildings, cats, cars, chairs, faces, fish, hammers, and phones. Data for cats, fish, hammers, and phones were not analyzed for the current study. All images of bodies, buildings, cars, chairs, and faces were the same images as those used in the behavioral masking experiment. Stimuli were presented in a rapid block design with each block corresponding to a particular category. Within a run, there were 90 total blocks with 10 blocks being devoted to each category. 6 images were presented per block for 667ms. The order in which the 90 blocks were presented was randomly determined at the beginning of each run. There were a random number of blank gaps in between blocks that could last 2, 4, or 6 seconds. The number and duration of the different gaps was determined at the beginning of each run using Optseq, which determines the optimal presentation of events for rapid-presentation fMRI (http://surfer.nmr.mgh.harvard.edu/optseq/). Participants were instructed to maintain fixation on a central cross and perform a vigilance task, pressing a button indicating when a red circle appeared around one of the images. The red circle appeared on 40% of blocks randomly on image 2, 3, 4, or 5 of that block.

Meridian map runs: Participants were instructed to maintain fixation and were shown blocks of flickering black and white checkerboard wedge stimuli, oriented along either the vertical or horizontal meridian (Sereno, et al., 1995; Wandell, 1999). The apex of each wedge was at fixation and the base extended to 8 degrees of visual angle in the periphery, with a width of 4.42 degrees. The checkerboard pattern flickered at 8 Hz. The run consisted of 4 vertical meridian and 4

horizontal meridian blocks. Each stimulus block was 12 s with a 12 s intervening blank period. The orientation of the stimuli (vertical vs. horizontal) alternated from one block to the other.

Localizer Runs: Participants performed a one-back repetition detection task with blocks of faces, bodies, scenes, objects, and scrambled objects. Stimuli in these runs were different from those in the experimental runs. Each run consisted of 10 stimulus blocks of 16 s, with intervening 12 s blank periods. Each category presented twice per run, with the order of the stimulus blocks counterbalanced in a mirror reverse manner (e.g. face, body, scene, object, scrambled, scrambled, objects, scene, body, face). Within a block, each item was presented for 1 s followed by a 0.33 s blank. Additionally, these localizer runs contained an orthogonal motion manipulation: In half of the blocks, the items were presented statically at fixation. In the remaining half of the blocks, items moved from the center of the screen towards either one of the four quadrants or along the horizontal and vertical meridians at 2.05 degrees per second. Each category was presented in a moving and stationary block.

fMRI Data Analysis

All fMRI data was processed using Brain Voyager QX software (Brain Innovation, Mastricht, Netherlands). Preprocessing steps included 3D motion correction, slice scan-time correction, temporal high-pass filtering (128 Hz cutoff), spatial smoothing (4mm FWHM Kernel), and transformation into Talairach space. Statistical analyses were based on the general linear model. All GLM analyses included box-car regressors for each stimulus block convolved with a gamma-function to approximate the idealized hemodynamic response. For each experimental protocol,

separate GLMs were computed for each participant and run, yielding beta maps for each condition.

Defining neural sectors

Sectors were defined in each participant using the following procedure. Using the localizer runs, a set of visually active voxels was defined based on the contrast of [Faces + Bodies + Scenes + Objects + Scrambled Objects] vs Rest (FDR<0.05, cluster threshold 150 contiguous 1x1x1 voxels) within a gray matter mask. To divide these visually-responsive voxels into sectors, the early visual sector included all active voxels within V1, V2, and V3, which were defined by hand on an inflated surface representation based on the horizontal vs. vertical contrasts of the meridian mapping experiment. The occipitotemporal and occipitoparietal sectors were then defined as all remaining active voxels (outside of early visual), where the division between the dorsal and ventral streams was drawn by hand in each participant based on anatomical landmarks and the spatial profile of active voxels along the surface. Finally, the occipitotemporal sector was divided into ventral and lateral sectors by hand using anatomical landmarks, specifically the occipitotemporal sulcus, which divides the ventral and lateral surfaces.

Group level brain/behavior correlation statistics (permutation analyses)

In order to test the significance of brain/behavior correlations at the group level, we performed a permutation analysis (Kriegeskorte, et al., 2008). To determine if a given correlation was significant, the condition labels of the data of each individual fMRI participant (N=6) and behavioral participant (N=20) were shuffled and then averaged together to make new, group

level structures. These structures were then correlated together and this procedure was repeated 10,000 times, resulting in a distribution of correlation values. A given brain/behavior correlation was considered significant if it fell within the top 5% of values in this distribution.

Linear mixed effects (LME) modeling:

Brain/behavior correlations in each sector, along with the differences in the strength of the correlations between sectors, were analyzed using a linear mixed effects model. To do this, the Fizher *z*-transformed correlation values were modeled as a function of the neural sector, including random effects of behavioral and neuroimaging participants on both the intercept and slope term of the model. The models were implemented using R (R Development Core Team, 2008) and the R packages *lme4* (Bates & Moechler, 2009) and *languageR* (Baayen, 2008). To determine if the correlations observed in the four sectors were statistically significant, we performed likelihood ratio tests, comparing a model with brain-region as a fixed-effect to another model without it, but which was otherwise identical including the same exact random effects structure. For significance testing, *P*-values were estimated using the normal approximation to the *t*-statistics (Barr, Levy, Scheepers, & Tily, 2013) and were considered significant if the *P*-values were below the a=0.05 value.

3

A ubiquitous and uniform representational structure across higher-level visual cortex

3.0 Abstract

Chapters 1 & 2 showed the relationship between behavioral capacity and high-level neural overlap. However, those studies were limited in their ability to identify the neural regions that were the most behaviorally relevant. Certain results from Chapter 1 (i.e. the lack of a difference in the strength of the correlations across occipitotemporal cortex and within category selective regions) suggest that the representational structure that is related to behavior may be widespread. In this Chapter, we first expanded the relationship between perceptual abilities and neural overlap using a new behavioral task: visual search. We then examined which neural structure best-predicted search performance. We found that virtually any higher-level neural structure correlated with behavior. These results suggest that the representational geometry in higher-level visual areas that constrains behavioral performance is highly uniform across higher-level visual cortex. The uniformity of this representational geometry, despite differences in response selectivity of regions like FFA and PPA, suggests that different regions of ventral visual cortex extract, and make explicit, different subsets of highly correlated perceptual features.

3.1 Introduction

Our visual systems are built to efficiently extract and encode the structure of the natural world, transforming the retinal input into intermediate and high-level representations of objects and scenes, which ultimately support our behavioral capacities (Olshausen and Field, 1996; DiCarlo & Cox, 2007). The representational structure of our visual system enables us to differentiate and identify thousands of different objects. A fundamental endeavor for visual cognitive neuroscience is to characterize this representational space of objects, discover how it is realized in neural substrates, and understand how it ultimately supports our behavioral object processing capacities (Ungerleider & Bell, 2011; Kourtzi and Connor, 2011; Cavanagh, 2011).

To relate neural processing with behavioral capacity, we adopted an analytical framework called representational similarity analysis (Kriegeskorte, et al., 2008). In this approach, the similarity between pairs of items/categories is calculated using a particular measurement (e.g. behavioral reaction times, neural correlations, etc.) with the set of these pairwise measurements coming together to form a *representational geometry* (Kriegeskorte & Kievit, 2013). This approach has been particularly useful in characterizing the similarity structure of neural responses, specifically in inferotemporal (IT) cortex (Kiani et al., 2007; Kriegeskorte, et al., 2008). In this case a number of representational dimensions have been discovered which predict the similarity of object responses and are embedded in large-scale spatial topographies: animate vs. inanimate (Kiani, et al., 2007; Kriegeskorte, et al., 2008).; Mahon & Caramazza, 2009; Konkle & Caramazza, 2013); big vs. small objects (Konkle & Oliva, 2012); faces, hands, bodies vs. other objects (Kanwisher, 2010). Finer-grained object distinctions are also present in the geometry, including distinctions between animals, buildings, and other object categories (Haxby et al.,

2001; Haxby et al., 2011; Huth et al., 2012). While this research has predominantly focused on characterizing the similarity structure of neural responses, the additional value of this approach is that representational geometries are easy to compare, not only across the neural regions (Kriegeskorte, et al., 2008; Cichy et al., 2014), but critically also between behavioral and neural measures (Op de Beeck et al., 2001; Carlson et al., 2013; Sripati & Olson 2010; Mur et al., 2013).

Here, we quantified the representational geometry of multiple high-level visual categories using both behavioral and neural measures. To get a behavioral index of similarity, we used a visual search task from which the similarity between any two object categories was calculated by measuring the time it takes to find a target from one category amongst distractors from another category (Duncan & Humphreys 1989). Meanwhile, to get a neural index of similarity, we used functional neuroimaging (fMRI) to measure and correlate the responses elicited by each of the categories when items were presented in isolation. We then asked if these different measures were correlated with one another, revealing a common representational geometry. To answer this, we independently analyzed different subsets of the visual hierarchy, which allowed us to examine a variety of hypotheses about the neural coding of these categories. For example, is the representational structure measured with visual search related to the structure found across the entirety of the dorsal and ventral visual pathways? Does this relationship change when only the well-known category selective regions for faces, bodies, and scenes are considered (Kanwisher, 2010)? Can any relationship be found when analyzing different fine-scaled structures within these pathways? Together, these analyses converge on the unexpected finding that the representational geometry measured by visual search is reflected in a ubiquitous and uniform fashion across higher-level visual cortex: it is present in both the dorsal and ventral streams (but

not early visual cortex), at a large scale, at a fine scale, and within and outside of category-selective regions. Taken together, these data suggest that our ability to discriminate between different categories is supported by a common, widespread coding space across both the ventral and dorsal streams.

3.2 Behavioral results

Participants (N=16) performed a visual search task that required discriminating between pairs of eight stimulus categories: bodies, buildings, cats, cars, chairs, faces, hammers, and phones (**Figure S1**). Stimuli were selected such that there would be a maximal amount of within-category diversity. For example, the body images were of people wearing different clothes, striking different poses, etc. In addition, all images were matched for average luminance, contrast, and spectral energy across the entire image (**Methods**). The selection and processing of the stimuli was done to minimize the extent to which visual search behavior could depend on low-level features (VanRullen, 2006).

The visual search task was comprised of eight blocks. Each block corresponded to a particular target category (i.e. "In this block, look for a face") (**Methods**). On every display, 8 items were presented equidistant from a central fixation mark (**Figure 12a**). Within each block, a target was shown on exactly half the trials, with the location of the target being randomly selected on a given trial. The critical manipulation was that each target presented display tested a particular category pairing (e.g. a target face amongst seven distractor chairs). Within each block, all of the distracting categories were presented the same number of times. With this design, reaction time measures were obtained for all of the 28 possible target-distractor category combinations.

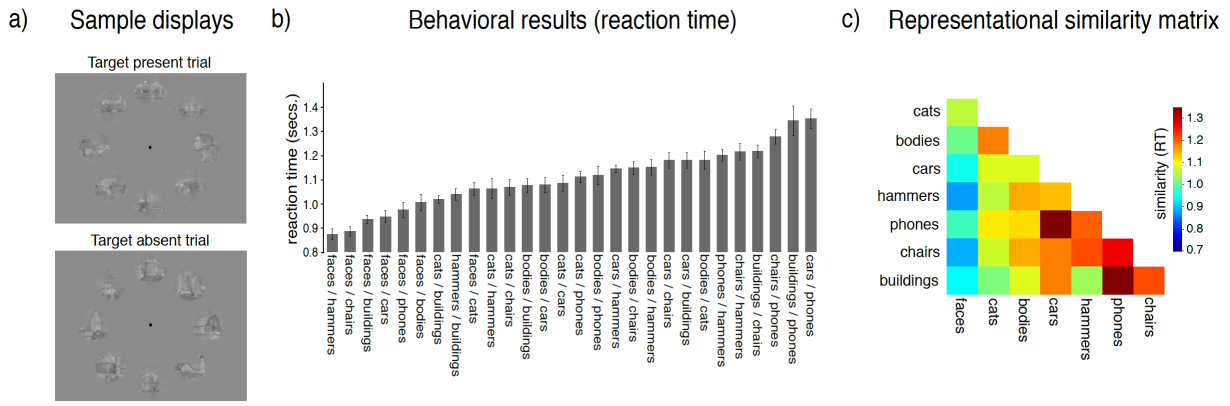

Figure 12: a) Sample displays from the visual search experiment. Top panel depicts a target present trial (a face amongst cars), while the bottom panel depicts a target absent trial (only buildings). b) Reaction time data from target present trials of visual search experiment when participants answered correctly. Reaction time (seconds) is plotted on the y-axis, with the different category pairings plotted on the x-axis. Note that each bar depicts the average reaction time for both target-distrctor combinations for a given category pairing (e.g. faces/hammers is both faces as the target and hammers as the target). Error bars represent the within-subject s.e.m. (Loftus & Masson, 1994). c) Representational similarity matrix based on the behavioral data. The dissimilarity measure is reaction time and is a re-plotting of the data from panel (b). In this case warmer colors denote slower reaction times/greater similarity, while cooler colors denote faster reaction time/less similarity.

To measure the representational geometry of this behavioral space, we focused on reaction times for target present trials in which the participant got the answer correct. In this case, the amount of time it takes to find a target from one category amongst distractors from another category indicates how similar those categories are in this particular space (Duncan & Humphreys, 1989). We focused on target present trials as they present the maximal opportunity for different categories to compete with one another for representational resources. On target present trials, not only has an internal template for the target category been formed that must be compared to every distracting item (Peelen & Kastner, 2011), the target and nearby distracting items on the display likely interact and compete with one another for neural resources (Desimone & Duncan, 1995). However, it should be noted that target-absent trials showed a

similar overall pattern of result, as the correlation between target-absent and target-present reaction times across the 28 category pairs was r=0.92 (see **Appendix C** for target-absent analysis).

Accuracy was near ceiling for each category pairing (**Methods**). Overall, we found significant differences in the time it took participants to find the target as a function of the target/distractor pairing ($F_{1,27}$=14.85, P<0.001) (**Figure 12b**), with visual reaction times varying by up to 465ms across the category pairings (faces & hammers: 887ms, cars & phones: 1,352ms). Several points can be made from a qualitative inspection of these data. First, category pairings in which faces were one of the two categories comprised the six fastest reaction times, with faces and cats being the 9th fastest. This suggests faces are relatively separated from the other categories in this particular similarity space (Hershler & Hochstein, 2006; Sinha et al., 2006) (**Figure 12c**). Second, category pairings that are within-animate (e.g. bodies and cats) or within-inanimate (e.g. cars and phones) are generally slower than pairings that are across animate-inanimate (e.g. cats and buildings). This is consistent with a variety of behavioral results showing that animacy is a primary psychological distinction (Caramazza & Shelton, 1998;) that is realized in the large-scale organization of the visual system (Martin, 2007; Kriegeskorte et al., 2008; Konkle & Caramazza, 2013).

3.3 Brain/behavior correlations

We then used fMRI to obtain whole-brain neural response patterns for each of these 8 categories in six new participants using a standard blocked design (**Experimental Procedures**). During scanning, images were presented in isolation and participants' task was to press a button

when a red circle appeared around an image. The neural patterns elicited by each category were correlated with one another across several cortical regions, yielding a representational geometry for each region (Kriegeskorte & Kievit, 2013). We then examined how these geometries relate to the one obtained from the visual search task by calculating a series of brain/behavior correlations. To do this, we considered a variety of neural hypotheses by evaluating different subdivisions of the cortex. First, we investigated whether the behavioral similarity structure is reflected across large swaths of cortex, which would support a large-scale coding hypothesis (Haxby et al., 2001; Kriegeskorte et al., 2008). This entailed focusing on the macro-scale divisions of the visual system: the ventral stream, the dorsal stream, and early visual cortex (V1-V3). Second, we explored the visual system at a finer spatial scale by using four methods to select subregions within the large-scale sectors: 1) category-selectivity, 2) overall responsiveness, 3) spatial proximity, and 4) random subsets of voxels. These analyses allow us to ask if a significant brain/behavior correlation can only be found within macro-scale sectors and also examine how the representational geometries at these finer spatial scales relate to each other and to behavior (e.g. do category selective regions such as FFA and PPA have different representational geometries and do they have different correlations with behavior?).

To preview the results, across all of these analyses targeting different neural coding hypotheses, we found significant brain/behavior correlations across the higher-level portions of both the ventral and dorsal streams at both the macro- and meso-scales. The fact that significant correlations were found within all meso-scale structures, including category selective regions, highlights a critical distinction between neural selectivity (i.e. tuning profiles) and neural representational geometry (i.e. the set of pairwise distances between items). We discuss the

implications of these findings for the neural architecture of object representation and its relationship to behavior.

3.3.1 Large-scale sectors

In each participant, macro-scale neural sectors were defined to capture early visual cortex, as well as the ventral and dorsal streams (**Figure 13a; Methods**). The early visual sector was defined based on retinotopic scans to functionally define areas V1-V3. Visually active voxels beyond these regions were subsequently divided into the ventral and dorsal streams based on functional activation profiles,yielding occipitoparietal and occipitotemporal sectors. Finally, given the mirrored representation of response selectivities across occipitotemporal cortex (Taylor & Downing 2011; Konkle & Carmazza, 2013), we further defined the ventral pathway into two ventral and lateral occipitotemporal sectors based on anatomical landmarks.

For each of these sectors, the correlations in the neural activation patterns between all category pairings were computed and were subsequently correlated with the behavioral similarity space based on reaction times (Kriegeskorte & Kievit, 2013). Statistical significance was assessed using a (i) group-level permutation analysis and (ii) linear mixed effects (LME) analysis. The first method is standard, but reflects fixed effects of both behavioral and neural measures, which is well suited for certain cases (e.g. in animal studies where the number of participants is small; Kriegeskorte et al., 2008). The second method is a more sensitive measure, given the power and number of participants in the current design, and uses a linear mixed effects model to estimate the brain/behavior correlation with random effects across both behavioral and neural measures (Barr et al., 2013; Winter, 2013; **Methods**). For visualization purposes, all figures show the

relationship between the group average neural measures and group average behavioral measures, with statistics from the group level permutation analysis for consistency. However, we also use the LME method to test the significance of the brain/behavior correlations and to compare the correlations between sectors, as this method captures the mixed within-subject (neural) between-subject (behavioral) nature of these comparisons.

We first assessed the relationship between the neural similarity structure of the higher-level sectors and the behavioral measure. We found strong brain/behavior correlations across both ventral stream subdivisions: ventral occipitotemporal (Permutation analysis: r=0.75, P<0.001; Mixed effects model: parameter estimate=0.64, t=11.81, P<0.001) lateral occipitotemporal (r=0.65, P<0.001; parameter estimate=0.49, t=6.78, P<0.001), and smaller, but still significant correlations in the dorsal stream sector: occipitoparietal (r=0.49, P<0.01; parameter estimate=0.26, t=3.15, P<0.05) (**Figure 13b**). However, the correlation in early visual cortex (V1-V3) was not significant (Permutation analysis: r=0.12, P=0.26; parameter estimate=0.09, t=0.97, P=0.33). This result was anticipated since the object exemplars were purposefully selected to have high variation and were matched in overall spectral energy (**Figure C.1**). Items were selected in this manner in order to put less emphasis on a retinotopic level of representation and instead focus on higher-level representations of complex features and object categories.

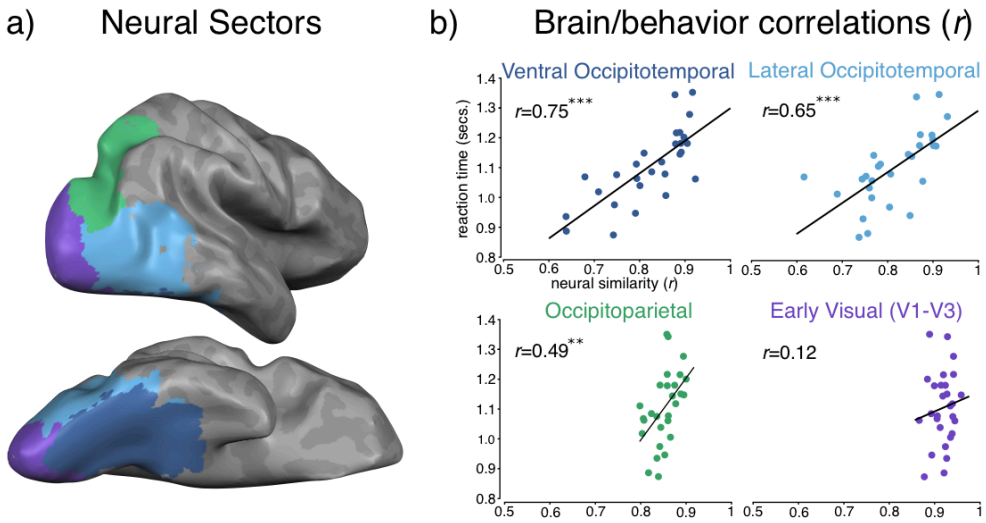

Figure 13: a) Visualization of the four neural sectors from a representative participant. b) Group level brain/behavior correlations in all sectors. Neural similarity between category pairings plotted on the y-axis. Estimated presentation durations for equal behavioral performance for all category pairings plotted on the x-axis. Each dot corresponds to a single category pairing (e.g. faces and buildings). The values on both axes were calculated by averaging across all behavioral and neuroimaging participants respectively and plotted accordingly. ***P<0.001, **P<0.01.

In addition, the relative strength of the brain/behavior relationships across these sectors was also reliable. That is, the brain/behavior correlation in ventral occipitotemporal cortex was significantly greater than those in all other sectors (parameter estimates <-0.15, t<-2.03, P<0.05 in all cases), the correlation in lateral occipitotemporal cortex was greater than those in both occipitoparietal and early visual cortex (parameter estimates <-0.23, t<-3.74, P<0.001 in both cases), and the correlation in occipitparietal was marginally greater than that in early visual cortex (parameter estimate=-0.16, t=-1.82, P=0.06).

These results suggest there is a strong relationship between perceptual and neural similarity: the neural similarity of these object categories across high-level visual cortex is strongly related to similarity as indexed by visual search. Given that the ventral stream sectorsencompass multiple category-selective regions (e.g. for faces/bodies/scenes), it is somewhat intuitive that these sectors would have the necessary response variation across

different object categories to capture the full behavioral similarity space. However, these results also provide support for the ventral/lateral distinction within the occipitotemporal cortex, as the ventral occipitotemporal cortex showed a reliably stronger relationship than lateral occipitotemporal cortex. While the difference between these two object streams is still unknown, the emerging representational distinction is between stimulus- (lateral occipitotemporal) and perceptual-bound (ventral occipitotemporal) representations (Haushofer et al., 2008; Taylor & Downing, 2011). Finally, while the dorsal stream is classically conceived of as representing spatial locations and coordinating visuo-motor actions (Ungerleider & Mishkin, 1982; Goodale & Milner, 1992), the correlations here strongly add to the mounting evidence that object category information is also encoded here (Sereno & Maunsell, 1998; Konen & Kastner 2008, Romero et al., 2014).

3.3.2 Category selective regions

Does a significant brain/behavior correlation require neural similarity to be pooled across large expanses of cortical territory, or is the same similarity structure present at finer spatial scales? To address this question, we tested whether or not functional subdivisions within these sectors also contain the requisite representational geometry to predict the behavioral similarity space.

First, we investigated the mosaic of category-selective regions for faces, bodies, and scenes (Kanwisher, 2010). On a strong modular account, the representational geometry of these regions could not correlate with behavior since these regions selectively process only one category and thus cannot capture the graded relationships between non-preferred categories (Spridon &

Kanwisher, 2002; Tsao et al., 2003; O'Toole & Haxby, 2005). However, it has been suggested that these regions do have some reliable discriminative responses to other categories in the finer-scale patterns, which suggests that more general object information is also present (Haxby et al., 2001; Reddy & Kanwisher, 2006). Critically, even on this latter account, it is not clear a priori whether the neural geometry of these object categories in a face-selective region (FFA) will be the same or different as those in a scene-selective region (PPA), and whether either of these will capture the full structure of the behavioral similarity space.

Regions preferring faces (FFA), scenes (PPA), bodies (EBA), and objects (LO) were independently localized in each participant (**Figure 14a, Appendix C)**, and the neural similarity for all category pairings was computed separately within each region. Across each of these regions, we observed strong correlations between the neural and behavioral similarity spaces: FFA (r=0.65, P<0.001; parameter estimate0.53, t=6.09, P<0.001), PPA (r=0.63, P<0.001; parameter estimate=0.52, t=6.01, P<0.001), EBA (r=0.56, P<0.01; parameter estimate0.34, t=4.18, P<0.001), and LO (r=0.49, P<0.01; parameter estimate=0.40, t=9.50, P<0.001) (**Figure 14b**).

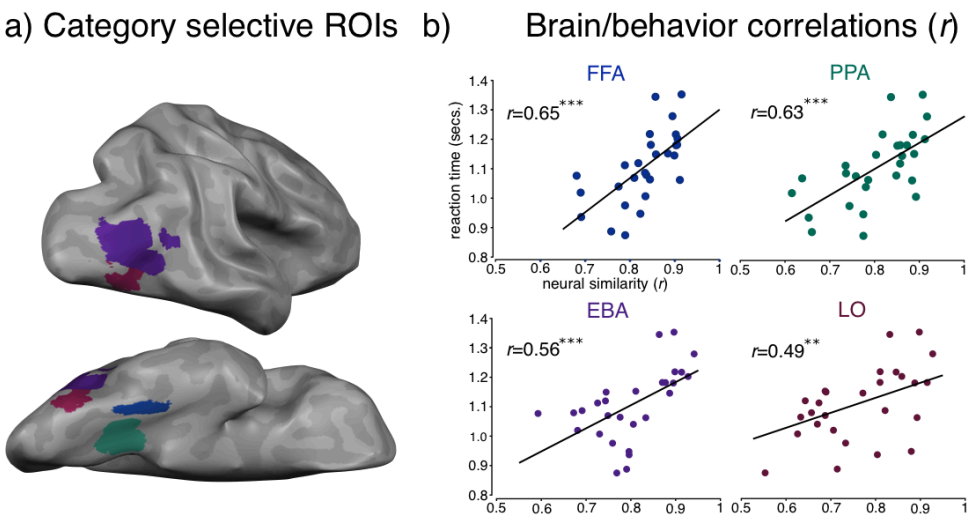

Figure 14: a) Visualization of the four category selective regions from a representative participant. b) Group level brain/behavior correlations in all sectors. Neural similarity between category pairings plotted on the x-axis. Estimated presentation durations for equal behavioral performance for all category pairings plotted on the y-axis. Each dot corresponds to a single category pairing (e.g. faces and buildings). The values on both axes were calculated by averaging across all behavioral and neuroimaging participants respectively and plotted accordingly. ***P<0.001, **P<0.01.

In comparing the strength of the correlation between these regions, we found a significant difference between FFA and LO (parameter estimate: -0.17, t=-3.85, P<0.001), with marginally significant differences between FFA and EBA, as well as PPA and EBA (parameter estimate>-0.18, t>-1.68, P<0.09 in both cases); all other pairwise comparisons were not significant (parameter estimate>-0.02, t>-0.15, P>0.40). As both FFA and PPA are on the ventral surface while EBA and LO are on the lateral surface of the human brain, these results are generally consistent with the previous macro-scale results showing that there are stronger ventral vs. lateral occipitotemporal brain/behavior correlations (**Figure 13b**).

It is important to consider whether the high brain/behavior correlations in these regions are driven entirely by responses to their preferred categories. For example, is it the case that in FFA, the representational distance between faces and the remaining categories (e.g. faces and bodies,

faces and cars, faces and cats, etc.) drives the correlation? To test this possibility, we removed all category pairings with faces from our analysis in FFA (leaving 21 object pairs for the brain/behavior correlation). In this case, we still found a significant correlation (r=0.67, P<0.001; parameter estimate=0.39, t=6.87, P<0.001) (see **Figure C.2**). Excluding faces and cats, which can elicit a strong response in the FFA (Tong et al., 2000), the correlation remained significant (r=0.61, P<0.001; parameter estimate=0.32, t=3.87, P<0.001). Further, even removing all animate categories (faces, cats, bodies)—leaving only the relationship between buildings, cars, chairs, hammers, and phones—the correlation between the FFA and behavior remained (r=0.54, P<0.01; parameter estimate=0.26, t=2.26, P<0.05). Similarly, the brain/behavior correlation was also maintained in the PPA when all pairings containing buildings were removed (r=0.64, P<0.001; parameter estimate=0.51, t=7.23, P<0.001), and in the EBA when bodies were removed (r=0.67, P<0.001; parameter estimate=0.57, t=4.49, P<0.001). These results clearly show that the preferred categories of these regions are not driving the brain/behavior correlations.

While each of the category-selective regions correlates with our behavioral measure, this result does not imply that they have the same underlying representational geometry; each region could be accounting for independent variance in the behavioral measure. To examine this possibility, we first directly compared the neural geometries across the category-selective regions to each other. In addition, we also analyzed the macro-scale sectors after removing all of the category selective regions. Surprisingly, we found that the neural similarity structure across all of the higher-level regions was significantly correlated (P<0.05 in all cases; **Figure 15**; **Figure C.3**). Meanwhile, the structure found within early visual cortex did not correlate with any other region

(P>0.11 in all cases). These analyses suggest that each of the category selective regions, as well as the higher-level macro-scale sectors, all have similar representational geometries.

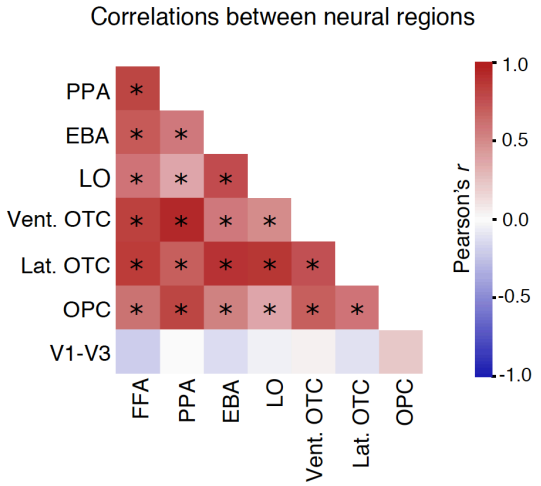

Figure 15: Correlations of the similarity structures between the category selective regions and macro-scale sectors when the category selective regions are removed. Thus, a region such as ventral occipitotemporal cortex (Vent. OTC) is all voxels in the sector when FFA, PPA, EBA, & LOC have been removed. The more saturated colors indicate stronger positive (red) and negative (blue) correlations. Asterisks within a cell indicates that the similarity structures between two given regions/sectors are statistically significant.

It is worth noting that these results highlight the fact that the tuning profiles of two regions can be very different, while their underlying representational similarity structure can be the same. **Figure 16** illustrates this for the FFA and the PPA. **Figure 16a** shows tuning profiles for the FFA and PPA across the 8 object categories. FFA and PPA have highly anticorrelated object preferences (**Figure 16b**). However, looking at the representational geometry of these regions defined by the neural similarity structure, it is evident that the pairwise-object similarities are highly related **(Figure 16c, d)**. While these results may appear at odds with each other, they actually both indicate the same representational geometry. Consider the responses to faces and buildings in FFA and PPA. FFA responds to faces far more than buildings, with the opposite

pattern occurring in PPA, as would be expected under a selectivity framework. However, from this opposite response pattern emerges the same conclusion about the relationship between faces and buildings in this representational space; namely, these categories are very far apart. Thus, even with dissimilar tuning profiles, it is still possible to find evidence for a common representational structure. These results are consistent with the idea that the FFA and PPA are a part of a common large-scale representational structure, where the maximal responses in each region can be interpreted as making explicit a particular part of this structure (as opposed to deriving a new representational space).

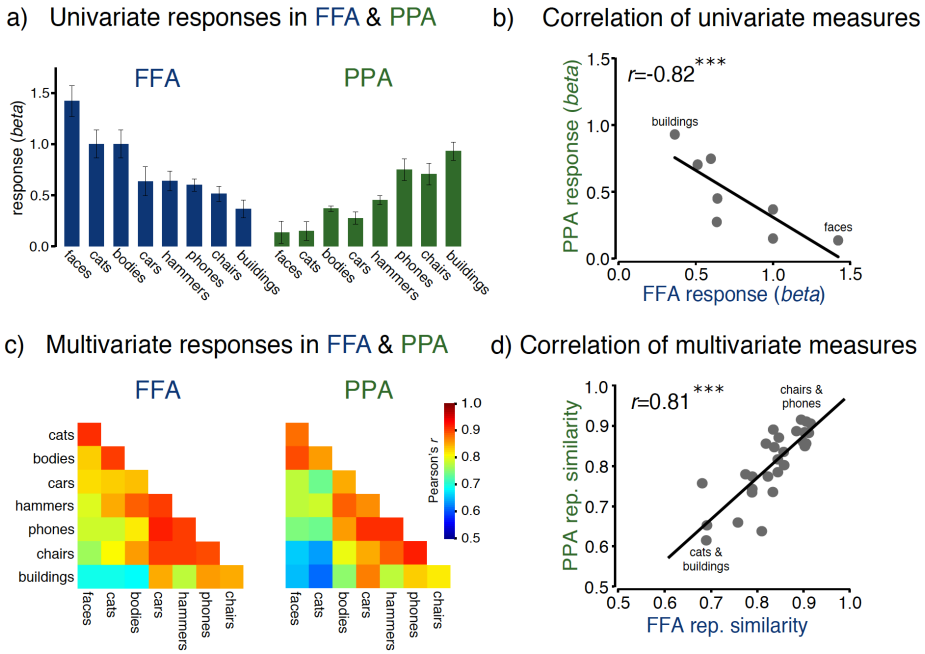

Figure 16: a) Univariate responses (beta) are plotted on the y-axis in FFA (blue) and PPA (green) for each of the eight categories, which are plotted on the y-axis. Error bars represent s.e.m. b) Correlation of the univariate measures from panel (a) between FFA (x-axis) and PPA (y-axis). Each dot represents the responses of a particular category (e.g. buildings or faces) in both FFA and PPA. d) Representational similarity matrix of the multivariate responses in both regions. The similarity measurement here is the correlation (*r*) of the patterns elicited by each category in both regions. Warmer colors denote higher correlations/similarity, while cooler colors denote lower correlations/similarity. d) Correlation of the multivariate measures from panel (c). PPA values are plotted on the y-axis and FFA values are plotted on the x-axis. Each dot represents the correlations between a particular category pairing (e.g. cats and buildings or chairs and phones) in both FFA (x-axis) and PPA (y-axis).

3.3.3 Alternative subdivisions

The analyses above demonstrated that the representational geometry measured by the behavioral task is also realized in meso-scale regions and does not require neural similarity measures that are pooled across large expanses of cortical territory. We next tested the robustness of this result using three alternative ways to subdivide the sectors into meso-scale regions.

First, we subdivided each sector based on its overall activation in response to all objects. Overall activity was calculated with an omnibus contrast of all object categories vs .rest, and this was used to divide each sector into 10 activation bins. For the occipitotemporal and occipitoparietal sectors, these subdivisions approximately reflected a posterior to anterior sweep across the cortical surface (see **Figure C.4** for visualizations of the most and least active bins). Within each activation bin, we calculated the neural similarity of the different category pairings and correlated this with the behavior similarity space. We found that the brain/behavior correlation was significant within each activation bin of both the ventral and lateral occipitotemporal cortex ($P<0.01$ in all cases) (**Figure 17a**). In occipitoparietal cortex, the brain behavior correlations were weaker, but were still significant across the top activation bins (bins 1-4: $P<0.05$ in all cases; bin 5 and 6 were marginal, $P=0.08$ and $P=0.07$ respectively). None of the activation bins in early visual cortex were significant ($P>0.10$ in all cases). Further, as was the case with the category-selective regions, all activation bins also contained highly correlated neural geometries (**Figure C.5**).

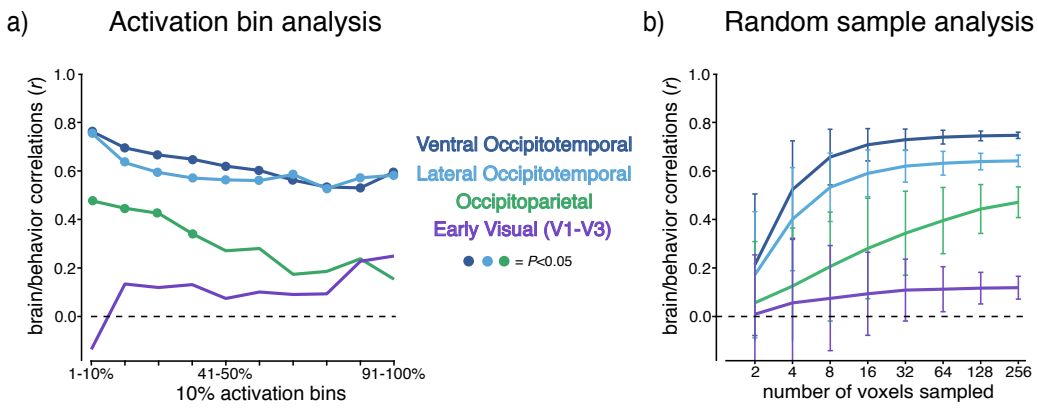

Figure 17: a) Activation bin analysis. The brain/behavior correlations between a given bin (e.g. Ventral OTC top 10% most active voxels) and the visual search data are plotted on the y-axis. The different bins are plotted on the x-axis. Filled circles indicate bins in which the brain behavior correlations were statistically significant. b) Random sample analysis. Brain/behavior correlations plotted (y-axis) as a function of the number of voxels randomly sampled 2,000 times (x-axis). Error bars represent the standard deviation of the distribution of correlation values for each sector at each sampling number.

Second, we explored the representational structure present in subsets of each sector by randomly sampling voxels. Following the procedure outlined in Hung et. al. (2005), voxels were randomly sampled from each sector and neural similarities across items in that sample were correlated with behavior. This process was repeated 2,000 times for each sample size (2 through 256, by powers of 2) within each sector. Results are shown in **Figure 17b**. Within ventral and lateral occipitotemporal cortex, the brain/behavior correlations steadily increased with sample size, reaching asymptote between 16-32 voxels. Overall, the results of this procedure demonstrate that this representational structure emerges with even a relatively small number of randomly scattered voxels.

Finally, we explored spatially contiguous subsets of voxels using a spherical searchlight analysis (Kriegeskorte et al., 2006). This analysis technique also permits us to investigate representational structure across the whole brain beyond our predefined visually responsive sectors (assuming a particular cortical scale defined by the sphere size). A sphere with a 3-voxel

radius was centered on each voxel, within which the neural similarity was computed across the category pairings and was then correlated with the behavioral measures (**Methods). Figure 18** shows the group-level brain/behavior correlations (*r*) on the inflated surface of a single subject. The results of this analysis generally converge with the previous results: 1) virtually every location within the ventral pathway is correlated with behavior; 2) approximately half of occipitoparietal cortex is significantly correlated with behavior, consistent with the activation bins analysis (**Figure 17a**); and 3) virtually no correlations are found in early visual cortex.

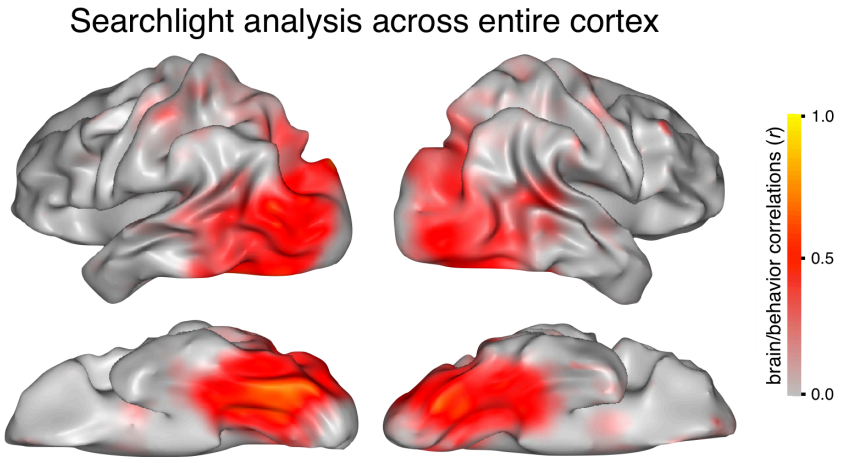

Figure 18: Whole-brain searchlight results for measuring the brain/behavior correlations (*r)* at each location of the cortex.

3.4 General discussion

Here we explored the extent to which a behavioral representational geometry is reflected in our neural architecture. We used a visual search paradigm to measure a behaviorally relevant similarity space of object categories, and used fMRI to obtain the neural response profiles for each category. Comparing the behavioral and neural measures across a variety of analyses yielded

three key results: (1) There is a striking correlation between the neural and behavioral similarity spaces across the macro-scale divisions of higher-level visual cortex that was not present early visual cortex. This relationship was particularly strong along the ventral surface of occipitotemporal cortex. (2) Across all methods of functionally subdividing the cortex into meso-scale regions, the neural similarity profiles across these regions were also highly correlated with behavior. (3) Surprisingly, the representational geometries of the higher-level macro- and meso-scale regions were correlated with each other.

These results reveal that the representational structure across the entire high-level visual cortex, along both the ventral and dorsal streams, at both macro- and meso-scales, has the requisite structure to predict the behavioral competition between object categories in visual search. Broadly, these results point to a ubiquitous and uniform representational structure in high-level visual cortex that supports and constrains our visual object processing capacities.

3.4.1 What is the nature of the representational space?

While these results demonstrate a clear relationship between object processing in behavior and high-level visual neural architecture, they do not directly reveal which particular features/dimensions of object categories comprise this structure. For example, this structure may be related to the visual appearance of object categories encoded from moderately complex image feature/fragments (Tanaka, 2003; Ullman, 2006), or related to more abstract attributes of object categories, such as their functional purpose (Mahon, 2007).

Discovering the nature of the features encoded by human occipitotemporal and monkey inferotemporal (IT) cortex is an extremely active topic (for a review see DiCarlo & Cox, 2007; Op

de Beeck et al., 2008; Ungerleider & Bell, 2011; Kourtzi & Connor, 2011; DiCarlo et al., 2012). To date, we know that retinotopic measures, such as pixel-wise and local-gabor functions, fail to capture the tolerance to visual transformations that is a signature of object responsive cortex (Kriegeskorte et al., 2008; DiCarlo et al., 2012). However, even current biologically inspired computational models built to capture more high-level visual features predict only part of the neural responses of IT cortex (Kiani et al., 2007; Kriegeskorte et al., 2008; Baldassi et al., 2013). These intermediate visual feature spaces typically fail to capture the category-level similarities among object exemplars. In contrast, pure semantic measures of object categories do capture some of the neural response variation, but these high-level spaces have too much categorical structure that is not graded enough to reflect the patterns in visual object cortex (Huth et al., 2012; Carlson, 2013; Mur et al., 2013).

Thus, the representational structure of this cortex seems to be right at the intersection between a perceptual and conceptual level, likely reflecting high-level visual shape features that are correlated with object categories. Our behavioral measure of visual search provides further evidence in support this general interpretation. For example, it is well-known that visual search is driven by a suite of visual features, from simple features like color and orientation (Wolfe & Horowitz, 2004), to more intermediate and complex shape features (Duncan & Humphreys, 1989; Suzuki, & Cavanagh, 1995), and correlates with explicit judgments of shape, color, and overall visual similarity (Konkle et al., 2010). However, visual search efficiency also depends on more abstract, experience-dependent factors, such as stimulus familiarity (Wang, et al., 1994; Malinowski & Hübner, 2001). These properties of our behavioral task complement the

computational approaches above, situating the likely level of representation as high-level, category-based shape features.

Importantly, our study was not designed to explore the nature of the representational dimensions of objects; instead it was designed to investigate how behavioral object processing capacities are related to the neural coding of object information. To this end, the macro-scale results indicates that there is are least two major representational spaces for visual information, one in early visual cortex presumably reflecting a more retinotopic level, and one across all of high-level visual cortex reflecting object-centered information. While the ventral surface of occipitotemporal cortex showed the strongest relationship with behavior, we cannot ignore that a strong relationship was also observed across the lateral occipitotemporal cortex and even in the dorsal stream sector of occipitoparietal cortex. Thus while we could have found evidence for more than two distinct levels of representation (e.g. separate coding spaces for the ventral, lateral, and parietal cortex), the work currently suggests that all of high-level cortex reflects a common coding space of high-level, category-based shape features. It will be important for future work to characterize the extent to which the ventral, lateral, and dorsal object streams represent distinct aspects of object processing information by using tasks and stimuli that accentuate these differences (Haushofer et al., 2008; Taylor & Downing, 2011; Harel, 2014).

3.4.2 Interpreting brain/behavior correlations

Here we showed that high-level visual cortex has the representational structure to predict a behavioral similarity space, but does this brain/behavior correlation isolate or explain the neural mechanisms that support behavior? For example, the neural responses measured here were in

response to objects in isolation while participants performed a simple vigilance detection task. However, when people perform a visual search task, neural responses shift towards the target object category. Recently, it has been shown that these attention-based tuning shifts are not only found in the category-selective regions of interest (Peelen & Kastner, 2011), but also over most if not all of object-responsive cortex (Cukur et al., 2013). These results are generally consistent with the idea that this cortex has a common representational geometry, but the question remains, what are the mechanisms by which we detect that a particular target, say a face among objects, is present in a display? Downstream regions may either read out the neural patterns over this whole cortex, or they may access a particular aspect of cortex where responses to faces over other objects drive the maximal neural responses. Each of these mechanisms requires different neural circuitry, and implies either a more distributed or a more localized code of object information. The present results indicate that similar information is available at both scales, and thus cannot isolate the behaviorally relevant level of representation. Critically, our goal was not to characterize the neural mechanisms of the search process. Instead, we aimed to characterize the representational space of objects over which search processes operate. To gain insight into the specific neural mechanisms contributing to behavior from this approach, future work should focus on tasks that reveal and accentuate different representational spaces (e.g. visual search, categorization, identification, etc.; Nakayama & Martini, 2011; Cox, 2014; Scalf et al., 2013).

3.4.3 Neural tuning vs. representational geometry

One of the more surprising results was that each of the category-selective areas correlated with the full behavior similarity structure, even when the preferred categories were excluded from the

analyses. These results highlight that there is an important distinction between neural tuning and representational geometry. That is, regions can have markedly different response preferences to some object categories over others (e.g. FFA and PPAs responses to faces and buildings), while at the same time retaining a nearly identical representational geometry across the set of object categories.

Much work has shown that tuning (i.e. relative response magnitudes) is important for neural representation. For example, causal studies using TMS and findings amongst patients with localized damage suggest that category-selective regions only play a role in processing the selected visual category (Farah, 2004; Epstein et al., 2001; Afraz et al., 2006; Pitcher et al., 2008; Dilks et al., 2013). These findings suggest that two regions with the same same geometry do not necessarily have the have the same functional role. This point is important in light of the recent expansion of studies using representational similarity methods to infer the function roles of cortical regions (Kriegeskorte & Kievit, 2013).

What, then, is implied by having a common representational geometry? We suggest that this points to a common level of representation across regions—that is, they respond to a common domain (visually-presented objects) with a common set of representational features (e.g. high-level category-relevant visual shape features). Regions with different selectivities can be thought of as making explicit different parts of a common representational space in a particular localized region (e.g. face-like objects, scene-like objects; Sato et al., 2013; Kornblith et al., 2013; Tanaka 2003). Such localized coding may be valuable as it supports neural mechanisms that can operate at a meso-scale (Chkovloskii & Koulakov, 2004; Kass, 1989), rather than requiring the read-out and manipulation of a much larger cortical territory. Indeed, this framing of object-responsive

cortex is analogous to that of early visual cortex: different neural regions explicitly code for particular locations in the visual field, with different orientation and spatial frequency preferences, while these regions are simultaneously part of a larger common representational space to encode the spatial-temporal structure of natural scene statistics across the visual field.

However, the present study focused on only eight categories, and this could have consequences for the interpretation of the relationship between selectivity and geometry. On one hand, these categories span the major divisions of objects that are known to date (e.g. animate body parts, big and small inanimate objects; Konkle & Caramazza, 2013). Thus, even with this limited number of categories, we are not likely to be limited by a restricted range when correlating neural similarity spaces. Moreover, with this limited set of categories, we were able to obtain highly reliable brain and behavioral measures that are essential for measuring such strong correlations (Nunnally, 1970). On the other hand, this power comes at the expense of testing more fine-grained relationships (e.g. within-category distinctions). An open possibility is that meso-scale regions may also have partially independent representational geometries reflecting more local representational subspaces, which are *not* found in other regions. Importantly, exploring finer-grained object variation across regions will necessarily have a more restricted range of neural response variation. Thus, even if these local representational geometries are identical across regions, the relationship might be difficult to detect due to the higher vulnerability to measurement noise with restricted range. Consequently, this question can only be addressed using high-fidelity neural measures in the context of explainable variance, before a reduced or absent correlation can be interpreted.

3.5 Conclusion

Here, we extend our understanding of the neural organization of object representation using two major approaches. First, we directly leverage behavior, using a behaviorally relevant representational space to investigate the representational structure underlying brain activation patterns. Second, we considered a variety of neural coding hypotheses, rather than assuming one particular spatial scale or functional division. In doing so, this reveled a ubiquitous and uniform representational structure across the ventral and dorsal streams, which is realized at both the macro- and meso-scales and is highly correlated with the behavioral representational structure of objects. These results in object-responsive cortex also raise important issues for the broader exploration of cortical representation, where the functional role of a region likely requires considering both the tuning profiles and the representational geometry present in the neural responses.

3.6 Methods

Stimuli

All stimuli are available for download at michaelacohen.wordpress.com/stimuli. The same eight stimulus categories were used for the behavioral and fMRI experiments: bodies, buildings, cars, cats, chairs, faces, hammers, and phones. Images were grayscaled and normalized on multiple low-level dimensions across the entirety of each image using the SHINE toolbox (Willenbockel et al., 2010). There were 30 exemplars within each category with each item being chosen such that there would be maximal within-category diversity (e.g. images of items taken from different angles, in different positions, etc.).

Behavioral Experiment Design

16 participants completed the behavioral experiment. The experiment was run on a 24-inch Apple iMac computer (1920 x 1200 pixels, 60 Hz) created and controlled with MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Participants sat approximately 57 cm away from the display so that 1 cm on the display would correspond to 1° of visual angle.

On each trial, 8 images were presented in a circular arrangement, 11.5° from a central fixation point (**Figure 12)**. The same 8 locations were used throughout the experiment. Each image was viewed through a flat top circular distribution and had a radius of ~3.25°. There were 8 experimental blocks, each with 10 practice trials and 112 experimental trials. Each block was defined by the target category (e.g. "Look for a face in this block."). Participants' task was to say whether or not a target was present on each trial. Responses were given via button presses on the keyboard with visual feedback given immediately. Within each block, a target was presented in a random location on half the trials. For each trial, the distractor items were selected from another object category. Across both target present and absent trials, each of the seven distracting categories appeared equally often. Within a block, all trial-types (e.g. target present/absent and which distracting items were shown) were presented in a random order. For each participant, trials in which the response time was either less than 300ms or greater than three standard deviations from the mean were excluded.

fMRI Data Acquisition

6 participants who did not complete the search task were collected on a 3T Siemens Trio scanner at the Harvard University Center for Brain Sciences. Structural data were obtained in

176 axial slices with 1 x 1 x 1 mm voxel resolution, TR = 2200 ms. Functional blood oxygenation level-dependent (BOLD) data were obtained using a gradient-echo echo-planar pulse sequence (33 axial slices parallel to the anterior commissure-posterior commissure line; 70 x 70 matrix; FoV = 256 x 256 mm; 3.1 x 3.1 x 3.1 mm voxel resolution; Gap thickness = 0.62 mm; TR = 2000 ms; TE = 60 ms; flip angle = 90 degrees). A 32-channel phased-array head coil was used. Stimuli were generated using the Psychophysics toolbox for MATLAB and displayed with an LCD projector onto a screen in the scanner that subjects viewed via a mirror attached to the head coil.

Experimental runs: Each participant completed 4 experimental runs, 1 meridian mapping run used to localize early visual cortex (V1-V3), and 2 localizer runs used for defining all regions of interest. Experimental Runs were part of a larger project within our laboratory. Thus, more categories were presented than were ultimately used for this project. Participants viewed images from nine categories: bodies, buildings, cars, cars, chairs, faces, fish, hammers, and phones. Fish were not presented in the visual search experiment (see above) and fMRI data from fish were not analyzed for this study. The same images from the search experiment were used in the fMRI experiment. Stimuli were presented in a rapid block design with each 4s block corresponding to one category. In each run, there were 90 total blocks with 10 blocks per category. 6 images were shown in each block for 667 ms/item. All runs started and ended with an 6s fixation block, and further periods of fixation that could last 2, 4, or 6 seconds were interleaved between stimulus blocks, constrained so that each run totaled 492s. The order of the stimulus blocks and the sequencing and duration of the fixation periods was determined using Optseq (http://surfer.nmr.mgh.harvard.edu/optseq/). Participants were instructed to maintain fixation on a central cross and perform a vigilance task, pressing a button indicating when a red circle

appeared around one of the images. The red circle appeared on 40% of blocks randomly on image 2, 3, 4, or 5 of that block.

Meridian map runs: Participants were instructed to maintain fixation and were shown blocks of flickering black and white checkerboard wedge stimuli, oriented along either the vertical or horizontal meridian (Sereno et al., 1995; Wandell, 1999). The apex of each wedge was at fixation and the base extended to 8° in the periphery, with a width of 4.42°. The checkerboard pattern flickered at 8 Hz. The run consisted of 4 vertical meridian and 4 horizontal meridian blocks. Each stimulus block was 12 s with a 12 s intervening blank period. The orientation of the stimuli (vertical vs. horizontal) alternated from one block to the other.

Localizer Runs: Participants performed a one-back repetition detection task with blocks of faces, bodies, scenes, objects, and scrambled objects. Stimuli in these runs were different from those in the experimental runs. Each run consisted of 10 stimulus blocks of 16s, with intervening 12s blank periods. Each category presented twice per run, with the order of the stimulus blocks counterbalanced in a mirror reverse manner (e.g. face, body, scene, object, scrambled, scrambled, objects, scene, body, face). Within a block, each item was presented for 1s followed by a 330ms blank. Additionally, these localizer runs contained an orthogonal motion manipulation: In half of the blocks, the items were presented statically at fixation. In the remaining half of the blocks, items moved from the center of the screen towards either one of the four quadrants or along the horizontal and vertical meridians at 2.05°/s. Each category was presented in a moving and stationary block.

fMRI Data Processing

All fMRI data was processed using Brain Voyager QX software (Brain Innovation, Mastricht, Netherlands). Preprocessing steps included 3D motion correction, slice scan-time correction, linear trend removal, temporal high-pass filtering (0.01 Hz cutoff), spatial smoothing (4mm FWHM Kernel), and transformation into Talairach space. Statistical analyses were based on the general linear model. All GLM analyses included box-car regressors for each stimulus block convolved with a gamma-function to approximate the idealized hemodynamic response. For each experimental protocol, separate GLMs were computed for each participant and run, which were then aggregated into an experiment-level multi-GLM, yielding beta maps for each condition.

Defining neural sectors

Sectors were defined in each participant using the following procedure. Using the localizer runs, a set of visually active voxels was defined based on the contrast of [Faces + Bodies + Scenes + Objects + Scrambled Objects] vs Rest (FDR<0.05, cluster threshold 150 contiguous 1x1x1 voxels) within a gray matter mask. To divide these visually-responsive voxels into sectors, the early visual sector included all active voxels within V1, V2, and V3, which were defined by hand on an inflated surface representation based on the horizontal vs. vertical contrasts of the meridian mapping experiment. The occipitotemporal and occipitoparietal sectors were then defined as all remaining active voxels (outside of early visual), where the division between the dorsal and ventral streams was drawn by hand in each participant based on anatomical landmarks and the spatial profile of active voxels along the surface. Finally, the occipitotemporal

sector was divided into ventral and lateral sectors by hand using anatomical landmarks, specifically the occipitotemporal sulcus.

To define category-selective regions (FFA/PPA/EBA/LO), we computed standard contrasts for face selectivity (faces >[bodies scenes objects]), scene selectivity (scenes >[bodies faces objects]), and body selectivity (bodies >[faces scenes objects]) based on independent localizer runs. For object-selective areas, the contrast of objects>scrambled was used. Category-selective regions included FFA (faces), PPA (scenes), EBA (bodies), and LO (objects). In each participant, face-, scene-, body-, and object-selective regions were defined using a semi-automated procedure that selects all significant voxels within a 9 mm radius spherical ROI around the weighted center of category-selective clusters (Peelen & Downing, 2005), where the cluster is selected based on proximity to the typical anatomical location of each region based on a meta analysis. All ROIs for all participants were verified by eye and adjusted if necessary. IAny voxels that fell in more than one ROI were manually inspected and assigned to one particular ROI, ensuring that there was no overlap between these ROIs.

Statistical Analyses and Significance Testing

Statistical significance was assessed using (i) group-level permuation analyses following Kriegeskorte et al. (2008) and (ii) linear mixed effects (LME) analysis. For the permutation analyses, the condition labels of the data of each individual fMRI and behavioral participant were shuffled and then averaged together to make new, group-level similarity matrices. The correlation between these matrices was computed and this procedure was repeated 10,000 times,

resulting in a distribution of correlation values. A given correlation was considered significant if it fell within the top 5% of values in this distribution.

For linear mixed effects modeling, we modeled the Fisher *z*-transformed correlation values between all possible fMRI and behavioral participants as a function of neural region. This included random effects analyses of neuroimaging and behavioral participants on both the slope term and intercept of the model, which was the maximal random-effects structure justified by the current design (Barr et al., 2013). All modeling was implemented using R (R Development Core Team, 2008) and the R packages *language* (Baayen, 2008) and *lme4* (Bates & Moechler, 2009). To determine if correlations were statistically significant, likelihood ratio tests were performed in which we compared a model with a given brain region as a fixed-effect to another model without it, but that was otherwise identical. *P*-values were estimated using a normal approximation of the *t*-statistic (Barr et al., 2013) and a correlation was considered significant if the *P*-values were below a=0.05

Searchlight

To examine the relationship between neural structure and behavioral performance within and beyond our selected ROIs, we conducted a searchlight analysis (Kriegeskorte et al., 2006). For each subject, we iteratively moved a sphere of voxels (3 voxel radius) across all locations within a subject-specific gray matter mask. At each location, we measured the response pattern elicited by each of the eight stimulus categories. Those patterns were correlated with one another, allowing us to measure the representational structure at each location. That structure was then

correlated with the behavioral measurements, resulting in an r-value for each sphere at every location.

4

Conclusion

4.1 General discussion

In this dissertation, we have argued that the extent to which different stimuli activate overlapping high-level neural channels, particularly those in occipitotemporal cortex, limit visual cognition. In Chapters 1 & 2, we showed that the amount of information that could be simultaneously encoded during a change detection task and the efficacy with which different stimuli masked other items from reaching awareness was predicted by measuring the extent to which the categories used in these tasks activate overlapping channels. In Chapter 3, we showed that this relationship between perceptual capacity and neural overlap also predicted visual search performance. In addition, we also found that there is a common, widespread representational structure that is related to these behaviors and can be found across the majority of higher-level visual cortex. Together, these results clearly demonstrate that a bottleneck of human cognition is the functional organization of the higher-level visual system.

One of the more surprising results from these studies is the fact that this measure of neural overlap successfully predicts all three behavioral tasks (i.e. change detection, visual masking, and visual search) even though each task places different demands on the observer. The change detection task requires that participants try to simultaneously encode several pieces of

information across space. The masking task, meanwhile, does not put pressure on the visual system by presenting items across space, but instead rapidly presents item across time. Finally, the visual search task requires that participants form internal templates of the target categories that will be compared to every considered item. The fact that strong brain/behavior correlations can be observed with each of these tasks suggests that this notion of channel overlap is a common bottleneck on visual cognition in general.

Of course, these results naturally beg one question: Are there any visual tasks will *not* be predicted by this neural measure within higher-level visual cortex? One obvious and likely answer is any task that relies upon low-level features (e.g. orientation, color, spatial frequency, etc.) that are not primarily represented in these higher-level cortical regions. For example, a visual search task for one target color (e.g. red) amongst differently colored distractors (e.g. orange) will likely not correlate with the amount of neural overlap associated with these colors in occipitotemporal cortex. Instead, those results would likely be predicted by lower-level parts of the visual hierarchy that explicitly code for color (e.g. V4). However, it should be stressed that this is not to say that a higher-level brain/behavior correlation will be observed whenever more complex, real world-stimuli are used. For instance, imagine a stimulus set in which each real-world category leaves a distinct "retinotopic fingerprints:" the bodies are all vertical and straight, the faces are all circular and forward facing, the cars are all horizontal and linear, etc. Even though the task instructions may be category based (e.g. "Look for a car in the display") the best predictor of the behavior may come from lower-level regions since observers may rely on these lower-level, retinotopic differences to perform the task. Thus, it appears that tasks in which

stimuli are relatively complex and cannot be discerned from lower-level similarities are the most likely to be predicted under this higher-level framework.

It should be stressed that this is not to suggest that performance on *any* task with high-level stimuli that are retiontopically variable (e.g. the stimuli used in Chs. 2 & 3 of this thesis) will correlate with overlap in occipitotemporal cortex. Indeed, certain behavioral paradigms seem to inherently rely on lower-level mechanisms and thus would not correlate with this neural measure. One example of such a task is binocular rivalry, the phenomenon in which two separate images are presented to both eyes and perception alternates between the images (Alais & Blake, 2005). Binocular rivalry is widely thought to arise from competition between relatively low-level mechanisms in the early visual hierarchy (Haynes & Rees, 2005; Tong et al., 2006, but see Logothetis, 1998). Therefore, it is unlikely that the rate at which two images alternate in rivalry would be predicted from higher-level neural measures even if the stimuli are selected to be sufficiently complex.

With these issues in mind, we suggest that the types of brain/behavior correlations observed here will be found when 1) the stimuli used are sufficiently complex and basic features (e.g. retinotopy or average luminance) cannot distinguish them, 2) when the task being used directly requires participants to encode the complexity of those stimuli, and 3) the neural mechanisms needed to complete the task are not restricted to early (e.g. occipital) cortex.

4.2 Future directions

Going forward, there are two natural research programs that follow from the work described here. One program entails explicitly detailing the scope and limits of the effects reported in this dissertation. For example, all of the reported results are based on analyzing both the behavioral and neural data at the *category level*. While there are many strengths to this approach (i.e. generalizability, statistical power, etc.), it does not reveal the extent to which this framework holds at the *exemplar level*. An easy way to examine this issue is to obtain reliable neural patterns in response to a large number of exemplars (~100) and acquire behavioral data for individual displays using the same types of tasks reported here. Such an experiment might reveal that the relevant representational structures will not be so widespread and uniform across higher-level visual cortex. Instead, the relevant neural networks may be far more localized. Such a result would highlight the boundary conditions of the effects reported here and if only more localized neural structures could predict certain behaviors, those regions could be studied more extensively. Indeed, such a brain/behavior correlation would suggest that that localized neural regions is the part of the visual system from which the relevant representations are being "read out" (Williams et al., 2007).

Another possible research program entails examining if the relationship between perceptual capacity and neural structure holds in other sensory modalities. For example, in the auditory modality, different types of stimuli (e.g. pitched sounds, speech, music, etc.) have been shown to activate distinct cortical regions and elicit unique activation patterns (Zatorre et al., 2002; Fedorenko et al., 2012; Norman-Haignere et al., 2013). Thus, it is possible that the capacity limits of auditory cognition will be predicted by the amount of overlap amongst these auditory

channels. One possible behavioral paradigm that naturally lends itself to this examination is a changes detection task similar to the one used in Chapter 1, but uses real-world auditory stimuli instead of visual images (Eramudugolla et al., 2005). If a similar pattern of results were found in the auditory modality, it would suggest that overlap amongst neural channels is a bottleneck on human cognition in general. Indeed, it may be the case that the organization of all sensory cortices — visual, auditory, tactile, etc. — may all create a functional/anatomical bottleneck on human cognition.

A

Appendix to Chapter 1:
Processing multiple visual items is limited by overlap in neural channels

A.1 Do outlier category pairings drive these brain/behavior correlations?

Given that there were only six category pairings used in this experiment, it is possible that one particular pairing (e.g. scenes & bodies) could be an outlier that drives the observed correlation. This possibility challenges our claim that the magnitude of the mixed-category benefit for any category pairing is strongly predicted by the amount of neural separation. To address this concern, we conducted two supplemental analyses.

First, we conducted a bootstrap analysis, in which we sampled, with replacement, 6 random categories pairings from the set, and conducted the same analysis on that random sample. Thus, on some iterations a particular category pair (e.g., scenes and bodies) will be left out completely, whereas on other iterations that pair could be included multiple times (giving more weight to that pairing). For each iteration, the random set of six category pairings were used to calculate a new, subject-averaged brain/behavior correlation. This process was repeated 10,000 times to obtain a distribution of correlation values. We found that in occipitotemporal cortex, a brain/behavior correlation of zero was not within the 95% interval when using the 10% overlap,

area under the curve, or pattern dissimilarity analysis ($P<0.05$ in all cases). As points of comparison, zero was within the 95% interval in occipitoparietal, early visual, and prefrontal cortex for all neural measure we used ($P>0.20$ in all cases).

Second, we also analyzed the data after having excluded each of the different category pairings one at a time (e.g. leave out faces & scenes and keep the rest, leave out faces & bodies and keep the rest). In this case, the brain/behavior correlations in occipitotemporal remained significant with all three neural measures regardless of which of category-pairings were left out ($P<0.01$ in all cases except when bodies & scenes were left out and we used the pattern dissimilarity measure, in which $P=0.59$). Taken together, these two results suggest that the significant brain/behavior correlations we find in occipitotemporal cortex are not driven by outlier category-pairings

A.2 Supplemental Figures

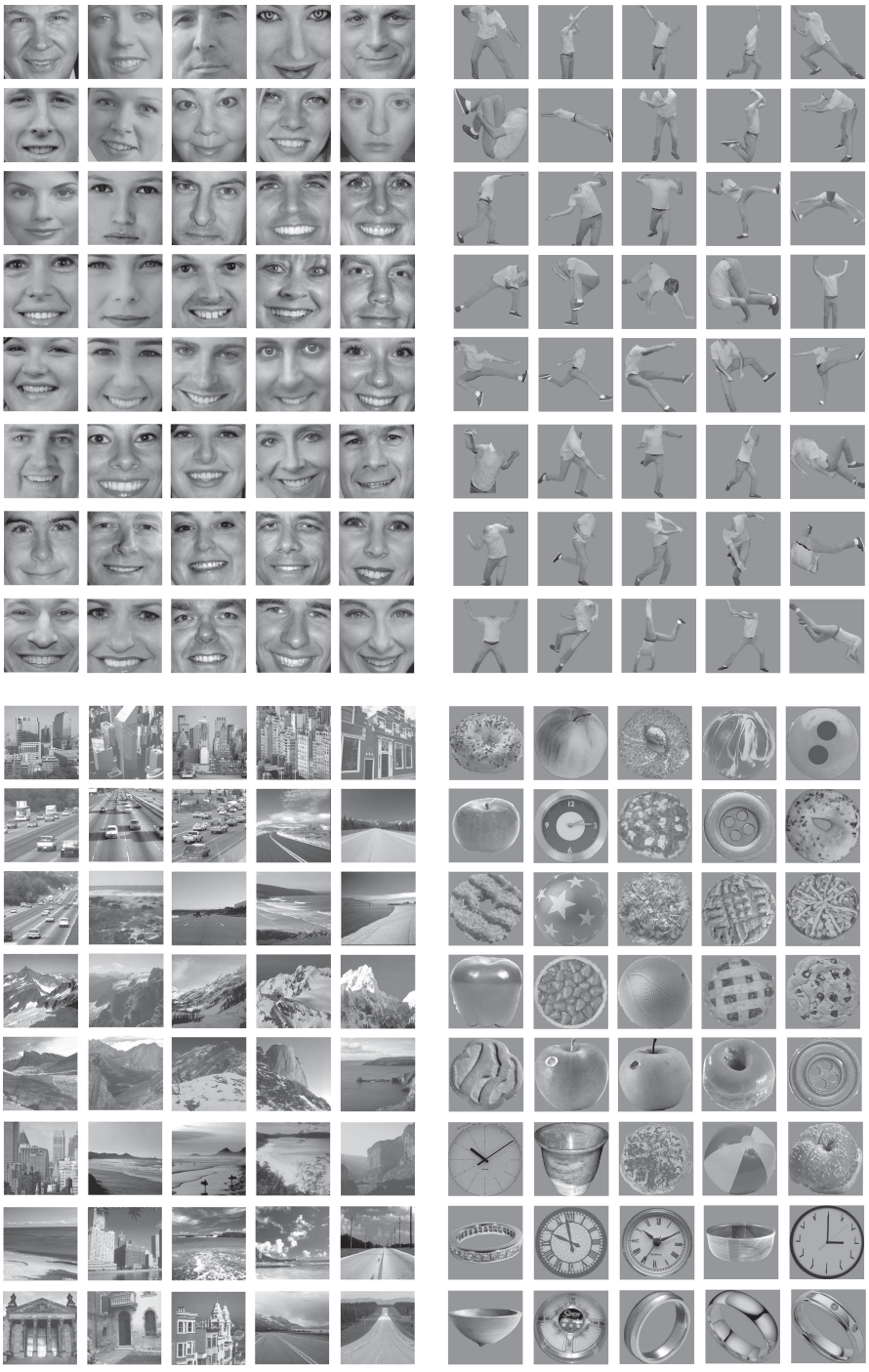

Figure A.1: Stimuli used in all experiments.

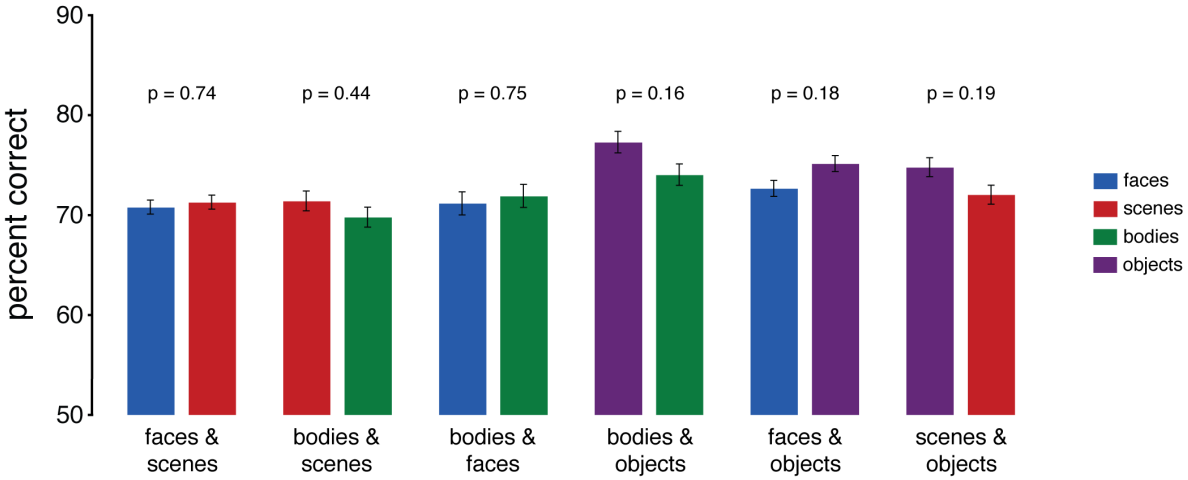

Figure A.2: Performance for same-category conditions across all category pairings. Each group of bars reflects one of the 6 possible category pairings (e.g. faces and scenes). Each bar reflects the percent correct on the displays in which all the items were from the same category (e.g. all 4 faces or all 4 scenes). Error bars reflect within-subject s.e.m. Note that for each category pairing, the data here were averaged together to compute the same-category performance data presented in **Figure 6.**

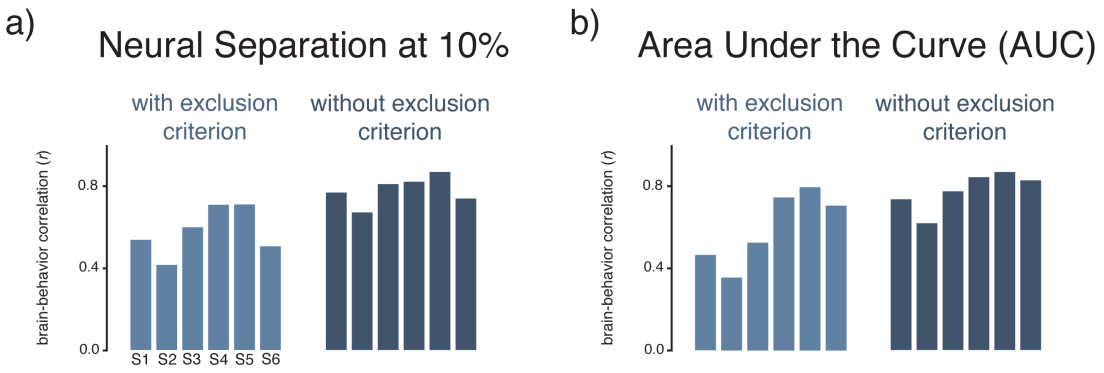

Figure A.3: Brain-behavior correlations in occipitotemporal cortex for each fMRI participant, with behavioral subjects excluded (light blue) and without behavioral subjects excluded (dark blue). The brain-behavior correlation (*r*) is plotted on the y-axis. Each bar represents an individual fMRI participant.

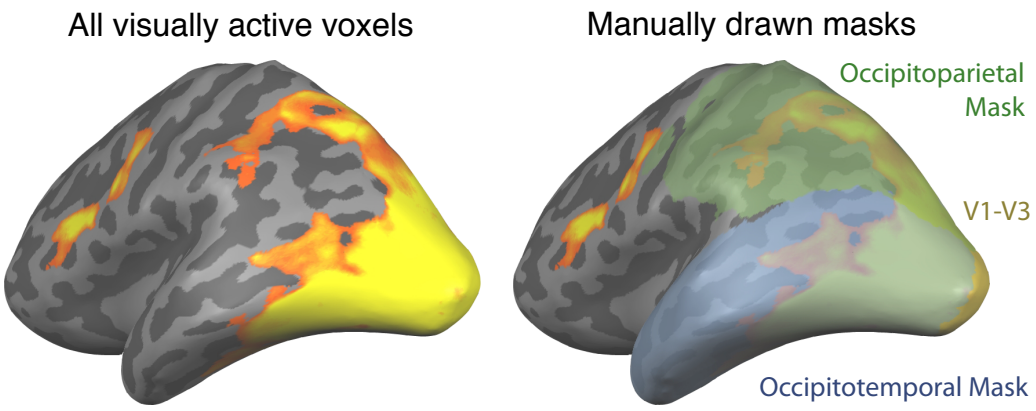

Figure A.4: All visually active voxels as obtained by the contrast [Faces+Bodies+Scenes +Objects] vs Rest (FDR<0.05, cluster threshold 150 contiguous 1x1x1 voxels) and shown on the inflated left hemisphere of a representative participant. For each participant, early visual cortex (V1-V3) was first defined with a Meridian Map on the inflated cortical surface. Once defined, masks for the occipitotemporal (light blue) and occipitoparietal (light green) cortices were manually drawn starting at the edge of V1-V3, up through the division between the ventral and dorsal pathways on the lateral surface, and continued to include all voxels within each pathway. It was from these masks that all active voxels were selected, which in turn because the ROIs used in the analysis.

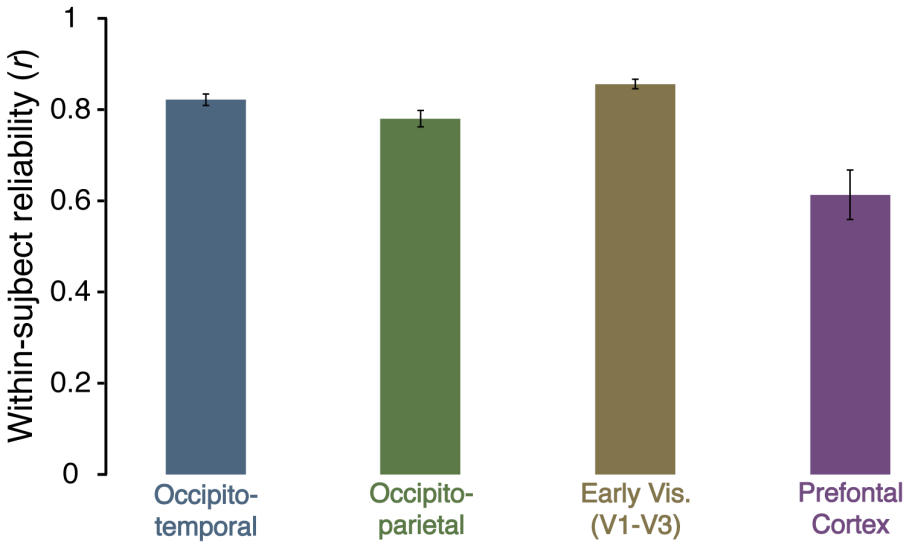

Figure A.5: Average within-subject split-half reliability of activation patterns, averaged across category, location, and subject. Error bars denote ± 1 standard error of the mean. See **Methods** for description of how within-subject reliability was calculated.

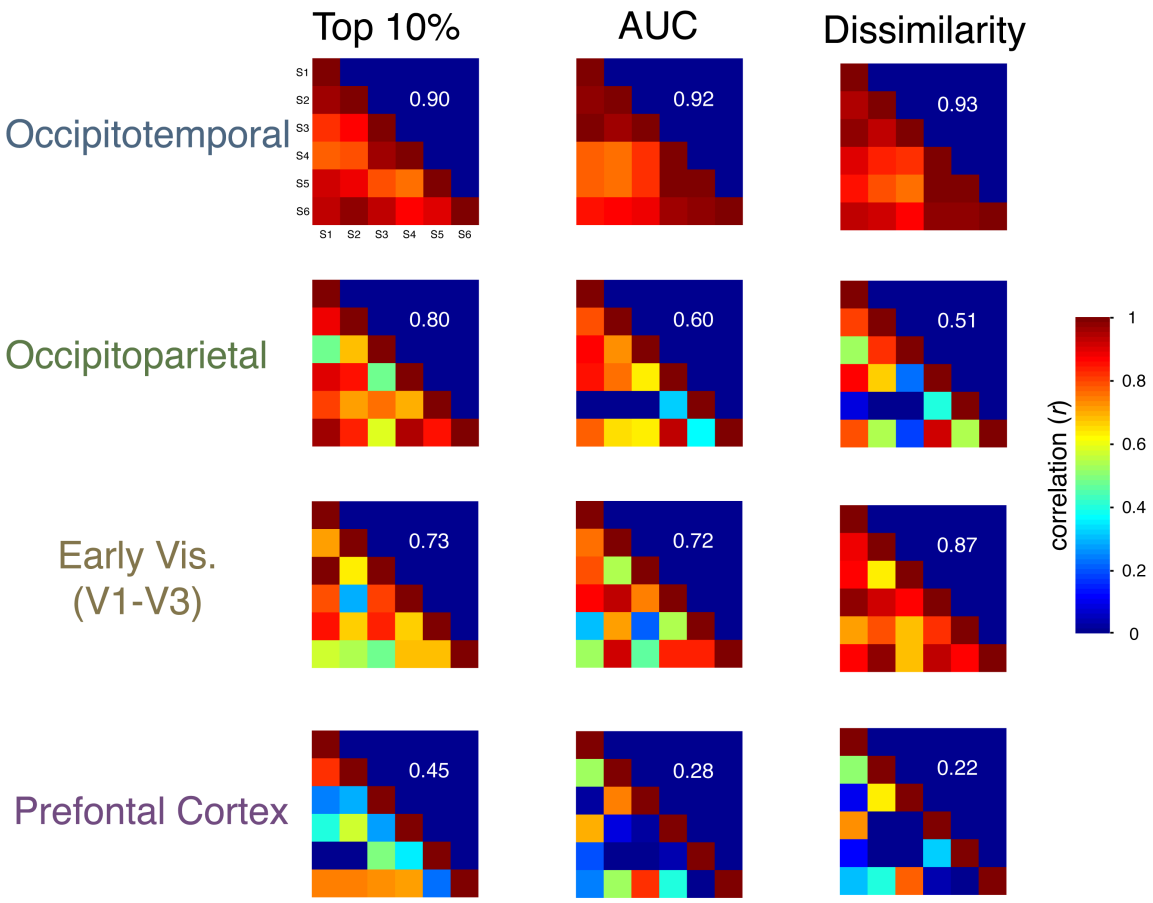

Figure A.6: Between-subject consistency in neural separation. Each plot shows the correlation between subjects in the amount of neural separation for each category pairing, with separate plots for each sector (rows) and for each neural separation measure (columns). Warmer colors denote higher correlations, while cooler colors denote lower correlations.

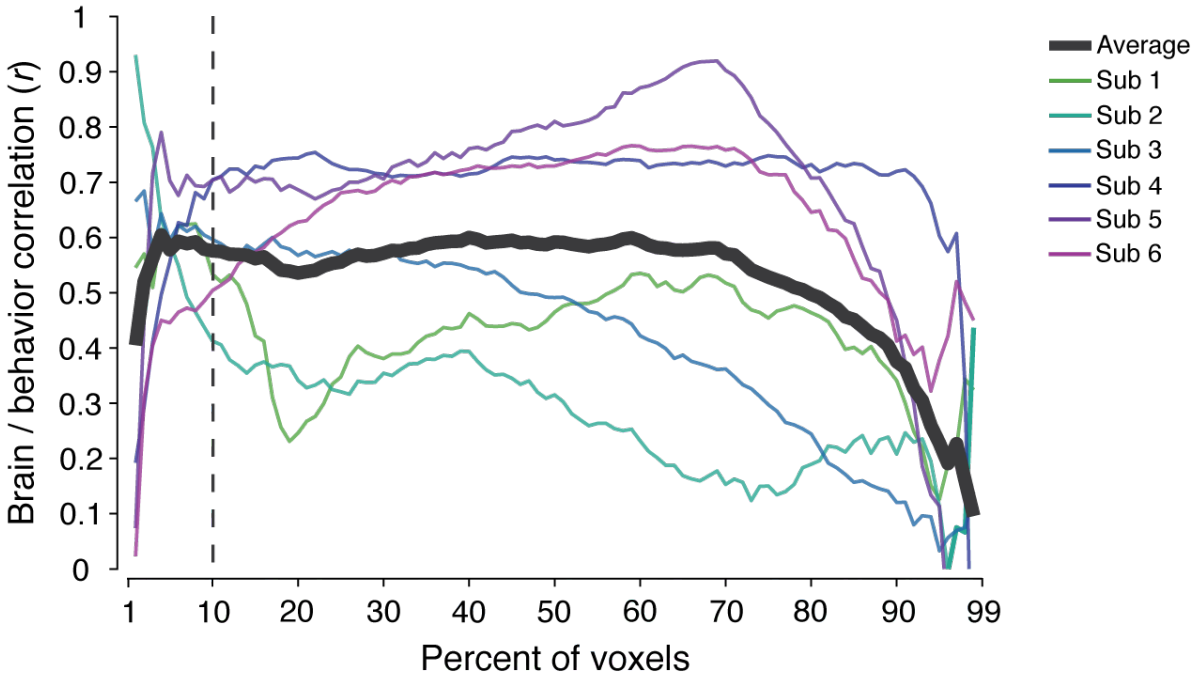

Figure A.7: Re-plotting of the brain/behavior correlation in occipitotemporal cortex from Fig. 5, showing the brain/behavior correlation as a function of activation cutoff**.** Individual fMRI participants are shown with each colored line and the group average is shown in the thick grey line. Percent of voxels considered for the overlap analysis are plotted on the x-axis and the brain/behavior correlation is plotted on the y-axis. The vertical dashed line indicates the 10% most active voxels (see **Figure 4**).

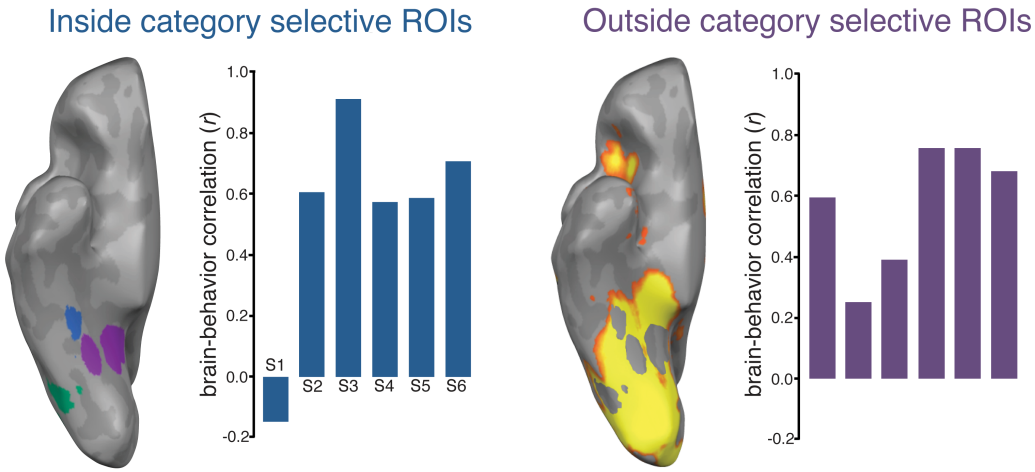

Figure A.8: Brain/behavior correlation considering the category-selective ROIs within occipitotemporal cortex (left panel), or excluding the category-selective ROIs within occipitotemporal cortex (right panel). For each category pairing, only the voxels selective for either of those two categories was included (faces: FFA/OFA; bodies: EBA/OBA; scenes: PPA/RSC; and objects: LOC). Each individual fMRI participant is plotted on the X-axis. The Y-axis shows the brain-behavior correlation (*r*) (see **Methods**).

B

Appendix to Chapter 2: *Overlap amongst visual processing channels limits visual awareness: Evidence from brain/behavior correlations*

B.1 Experiment 1 Supplement

A natural question to ask is if there were ever instances in which the two QUEST estimates produced within a given block were significantly different from one another. Indeed, there were multiple cases in which one target/mask combination had a significantly different presentation estimate than the opposite target/mask combination. For example, the combination of face target/building mask had an estimate of 52ms, while the combination of building target/face mask had an estimate of 29 ms. All estimates can be seen in **Figure B.1**.

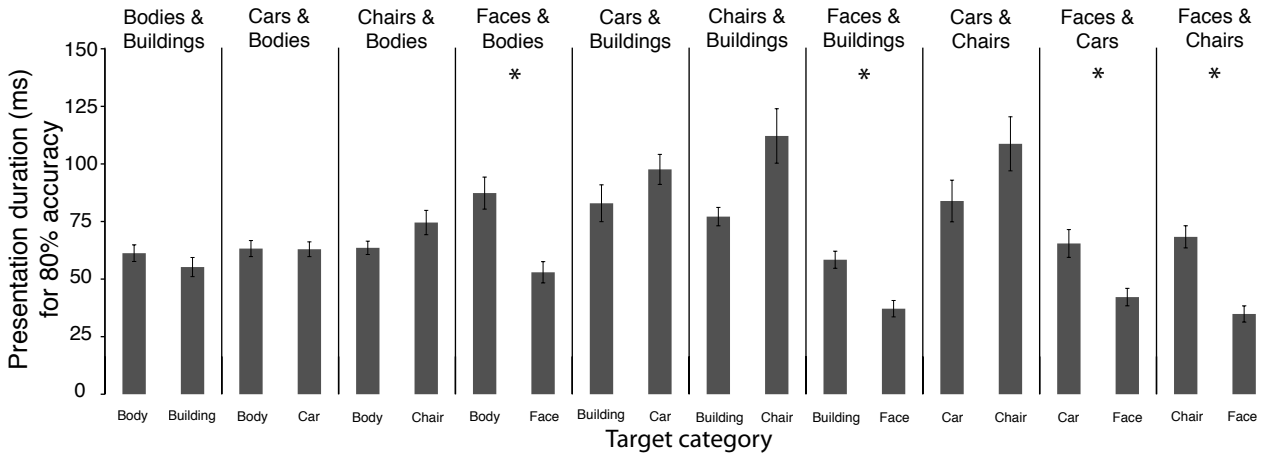

Figure B.1: Behavioral results from Experiment 1. Estimated presentation rates needed to result in equal performance for each target/mask configuration are shown on the y-axis. Target categories are shown on the x-axis. Thus, for the faces and buildings pairing, the bar marked "Building" reflects the estimated presentation rate when buildings are the targets and faces are the masks. Error bars reflect within-subject s.e.m. * denotes $P<0.05$ Bonferroni corrected for multiple comparisons.

Asymmetries in the threshold duration for a particular category pairing could reflect two very different sources. First, it is possible that such asymmetries simply reflect an attentional bias (e.g., participants might default to looking for faces). By averaging across the particular target/mask configuration (e.g., averaging estimates for face among bodies with body among faces), we can factor out such attentional biases. However, it is also possible that these asymmetries reflect asymmetries in the similarity between categories (e.g., Tversky, 1977). One way to address this question is to develop asymmetric measures of neural similarity (e.g., akin to the Tversky index). However, this approach is beyond the scope of the current paper, and for the present purposes, we simply investigate whether the observed brain-behavior correlation patterns change when these differences are taken into account? To examine this we plotted the behavioral data from Experiment 1 after breaking the data down by the different target/mask configurations and compare those values to the neural measurements. In this case, we plot the two possible

target/masking configurations from a single category pairing (e.g. faces & bodies) against a single neural similarity value obtained for that category. Since each neural value is plotted against two behavioral values, the points in the scatter plot are not independent of one another. Thus, standard significance testing of correlations is not appropriate since such tests assume independence amongst all measurements. For this reason, we performed a permutation analysis in which the labels of the different conditions for both the behavioral and neural data were shuffled and a correlation between those shuffled values were correlated with one another. This process was repeated 10,000 times to obtain a distribution of correlation values. An observed brain/behavior correlation was considered statistically significant if it fell within the top 5% of that distribution (Kriegeskorte, Mur, Ruff, Kiani, Bodurka, Esteky, Tanaka, Bandettini, 2008).

As was the case in the main analysis, we found strong correlations in ventral occipitotemporal cortex (r=0.69, P<0.001), with lesser correlations in lateral occipitotemporal and occipitoparietal cortex (r=0.58, P<0.01 and r=0.43, P<0.05 respectively), with no correlation in early visual cortex (r=0.04, P=0.42) (**Figure B.2**). It is worth noting that the correlations in the three higher-level cortices are higher when the behavioral data are averaged together compared to when it is not. This is likely due to the fact that in in splitting each behavioral measure into two distinct measurements, each measurement has half as many trials and is thus a noisier measure, which would decrease the correlations with the neural data.

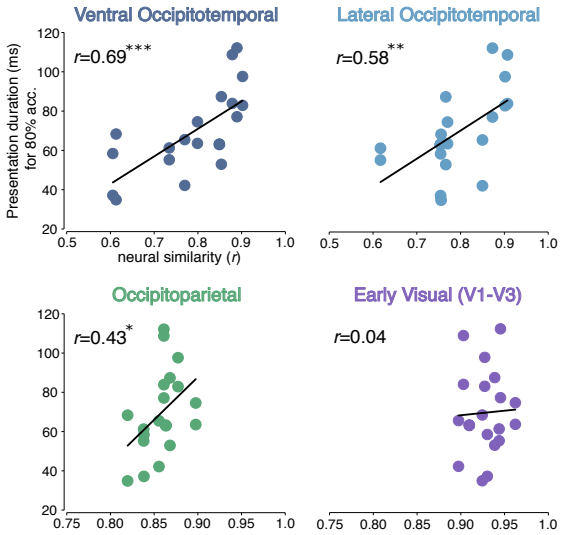

Figure B.2: Group level brain-behavior correlations in all sectors in Experiment 1 when the behavioral data is broken down by each target/distractor combination. Neural similarity between category pairings plotted on the y-axis. Estimated presentation durations for equal behavioral performance for all category pairings plotted on the x-axis. Each dot corresponds to a single target/distractor pairing (e.g. face target and building masks). Note the change in the scales of the x-axis between the two occipitotemporal and occopitoparietal/early visual plots. ***P<0.001; **P<0.001; *P<0.05

B.2 Experiment 2 Supplement

Results from Experiment 2, in which the different QUEST estimates were averaged together, can be seen in **Figure B.3**. The correlation between this data and the same data obtain in Experiment 1 can be seen in **Figure B.4**.

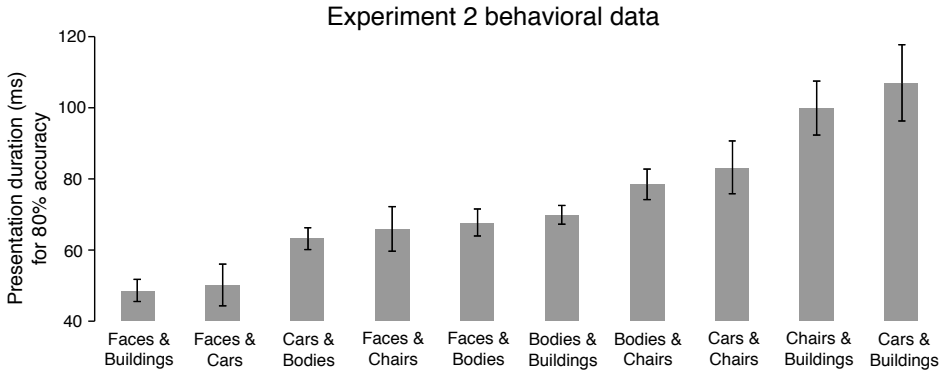

Figure B.3: Behavioral results from Experiment 2. Estimated presentation rates needed to result in equal performance for each category pairing are shown on the y-axis. Each bar corresponds to a particular category pairing. Error bars reflect within-subject s.e.m.

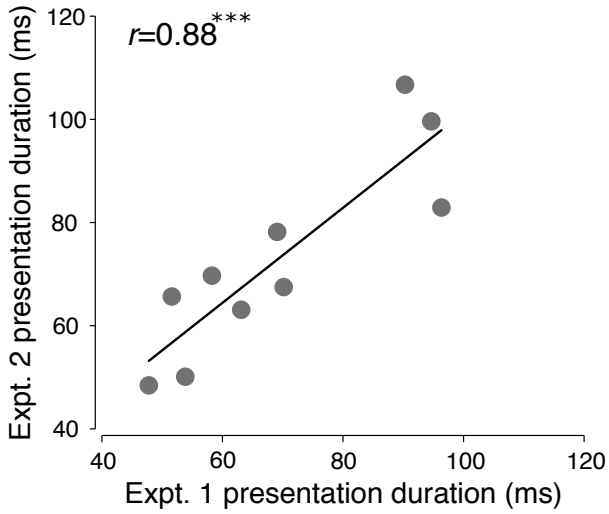

Figure B.4: Correlation between behavioral data from Experiments 1 and 2. Each dot corresponds to a given category pairing.

Once again we asked how the pattern of brain-behavior correlations changes when the data from Experiment 2 is not averaged together. As was the case with Experiment 1, there were instances in which the estimates from within a category pairing were significantly different from one another (**Figure B.5**).

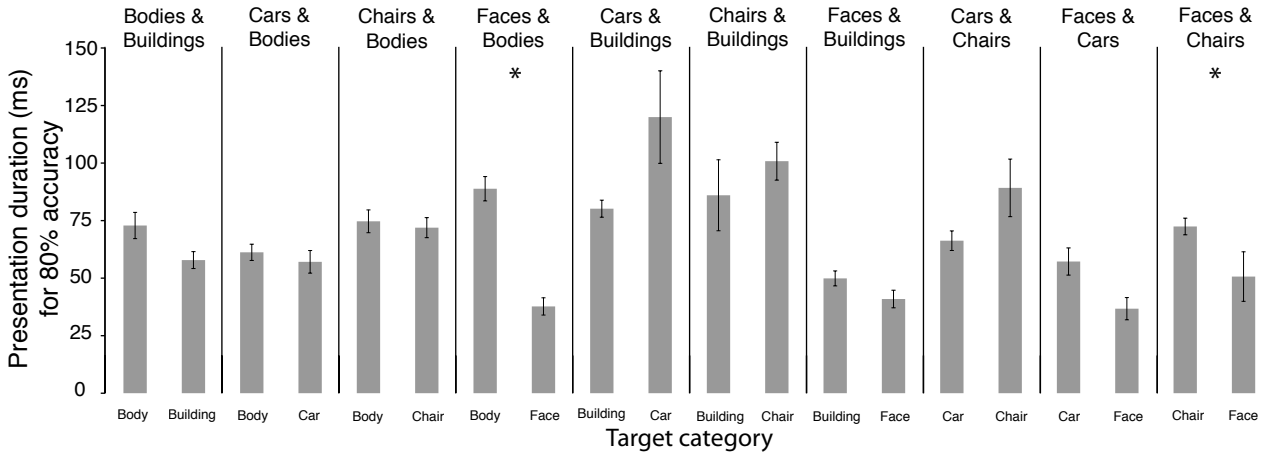

Figure B.5: Behavioral results from Experiment 2. Estimated presentation rates needed to result in equal performance for each target/mask configuration are shown on the y-axis. Target categories are shown on the x-axis. Error bars reflect within-subject s.e.m. * $P<0.05$ Bonferroni corrected for multiple comparisons.

In addition, for each category pairing, we calculated the difference in the QUEST estimates for both target/distractor combinations (i.e. a 16ms difference between body target and buildings mask compared to building target and bodies mask. See **Figure B.5** left-most two columns). We did this for both Experiment 1 and 2, and then correlated the difference values between the two experiments with one another. In this case, we found a strong correlation between these values (r=0.84, P<0.01). The fact that we observed strong correlations in the behavioral estimates for Experiments 1 and 2 when the data was (**Figure S4**) and was not averaged together further demonstrates that the process of averaging the data together for the main analysis does not change the main pattern of results.

As we did in Experiment 1 (**Figure B.2**), we once again broke the behavioral data down by every possible target/distractor combination. Once again we found strong correlations in ventral occipitotemporal cortex (r=0.56, P<0.01), with lesser correlations in lateral occipitotemporal and occipitoparietal cortex (r=0.41, P<0.05 and r=0.37, P=0.05 respectively), with no significant correlation in early visual cortex (r=0.25, P=0.14) (**Figure B.6**).

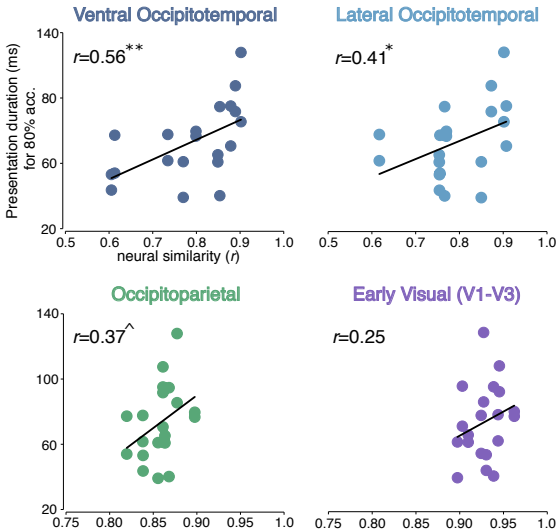

Figure B.6: Group level brain-behavior correlations in all sectors in Experiment 2 when the behavioral data is broken down by each target/distractor combination. Neural similarity between category pairings plotted on the y-axis. Estimated presentation durations for equal behavioral performance for all category pairings plotted on the x-axis. Each dot corresponds to a single target/distractor pairing (e.g. face target and building masks). Note the change in the scales of the x-axis between the two occipitotemporal and occopitoparietal/early visual plots. ** $P<0.001$; * $P<0.05$, ^ $P<0.10$.

C

Appendix to Chapter 3:
A ubiquitous and uniform representational structure across higher-level visual cortex

C.1 Target present vs. target absent analysis

In the main text, we focused our analyses on target present trials; How does the pattern of results change, if at all, on target absent trials? Since the two sets of behavioral data were highly correlated with one another (r=0.92, P<0.001), we predicted that the relationship between neural structure and behavioral capacity would not vary depending on which behavioral data was analyzed. Indeed, we found that the same basic pattern of results was found with target absent trials as was found with target present trails. Together with the results reported in the main text, these results suggest that the primary level of competition between items occurs between an internal template for each target and the distractor items being processed. See below for brain/behavior correlation analyses for target absent trials.

Brain/behavior correlations in macro-scale sectors and category selective regions:

When using target absent trials to estimate the behavioral similarity space, significant brain/behavior correlations were observed in each higher-level macro-scale sector: ventral

occipitotemporal (r=0.78, P<0.001; parameter estimate=0.73, t=10.74, P<0.001), lateral occipitotemporal (r=0.61, P<0.001; parameter estimate=0.53, t=6.13, P<0.001), and occipitoparietal cortex (r=0.56, P<0.001; parameter estimate=0.30, t=4.25, P<0.001), but not in early visual cortex (r=0.16, P=0.22; parameter estimate=0.08, t=1.03, P=0.30).

In addition, significant correlations were found in every category selective region (FFA: r=0.62, P<0.001, parameter estimate=0.55, t=5.96, P<0.001; PPA: r=0.70, P<0.001, parameter estimate=0.66, t=5.64, P<0.001; EBA: r=0.58, P<0.01, parameter estimate=0.38, t=3.94, P<0.001; LO: r=0.41, P<0.05, parameter estimate=0.32, t=4.13, P<0.001).

10% Activation Windows and Random Sampling

When calculating the brain/behavior correlations in the 10 activation bins in each macro-level sector, we once again found significant brain/behavior correlations in each bin in both ventral and lateral occipitotemporal cortex (P<0.01 in all cases), as well as in the first 5 bins in occipitoparietal cortex (P<0.05 in all cases) with the 6th being marginally significant (P=0.07). In addition, we also performed the random sample analysis in which between 2-256 voxels were randomly sampled from each sector and compared to the target absent data. As was the case with the target present data, we found that the brain/behavior correlations in the occipitotemporal sectors when considering target absent data also plateaued between 16-32 voxels (**Figure S4**).

C.2 Supplemental Figures

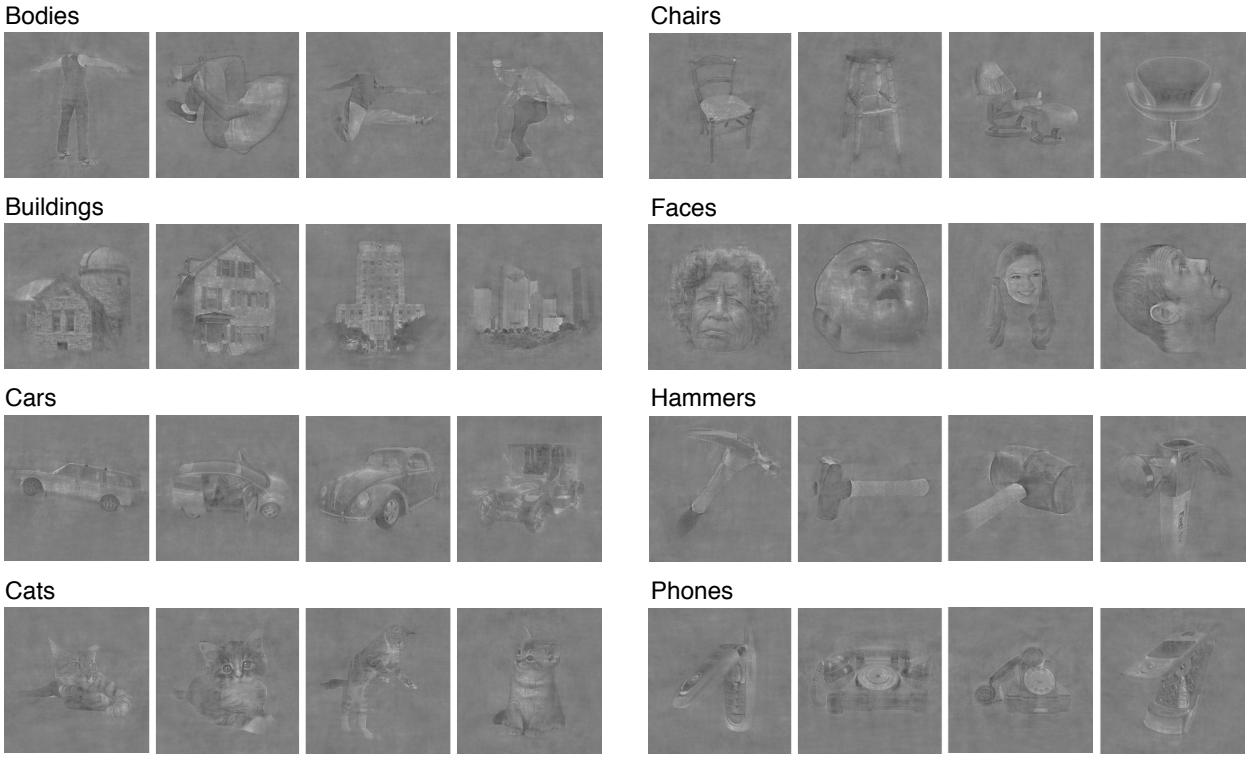

Figure C.1: Examples of stimuli from each of the 8 categories.

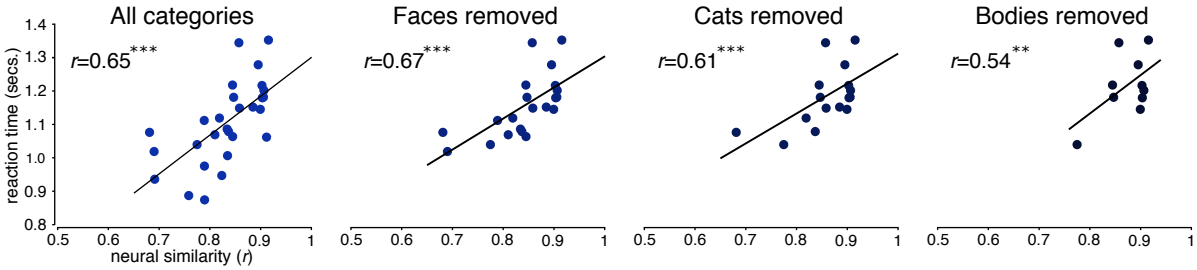

Figure C.2: Brain/behavior correlations within FFA as a function of which category pairings are considered. Reaction time data from the visual search task is plotted on the y-axis, with neural similarity values between the category pairings in FFA plotted on the x-axis. From left to right, brain/behavior correlations when all categories are considered, when faces are removed, when cats are removed, and when bodies are removed. ***P<0.001, **P<0.01.

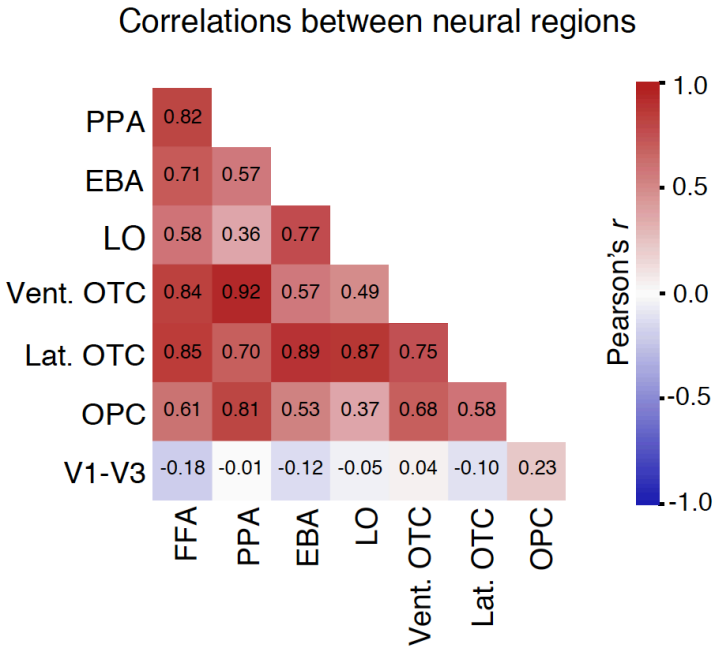

Figure C.3: Correlations of the similarity structures between the category selective regions and macro-scale sectors when the category selective regions are removed. This is the same data as plotted in **Figure 15** with the *r*-value filled in for each cell. Note that all correlations are statistically significant except those that include V1-V3 (i.e. the bottom row).

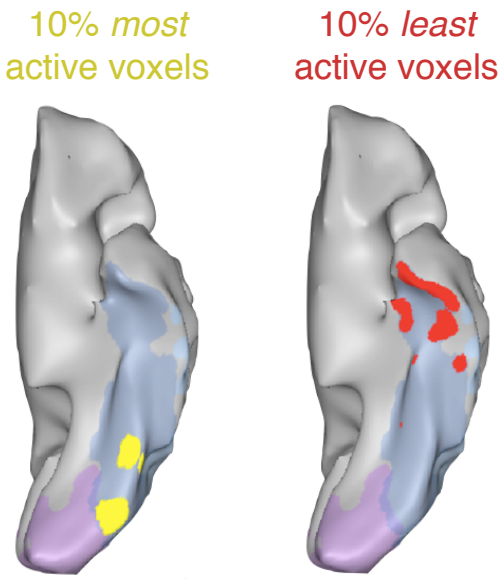

Figure C.4: The top (left hemisphere/yellow) and bottom 10% (right column/red) activation bins in ventral occipitotemporal cortex of a representative subject. Note how there are a few contiguous clusters of voxels that together comprise both bins and the difference in their posterior vs. anterior location for top vs. bottom 10% activation bins.

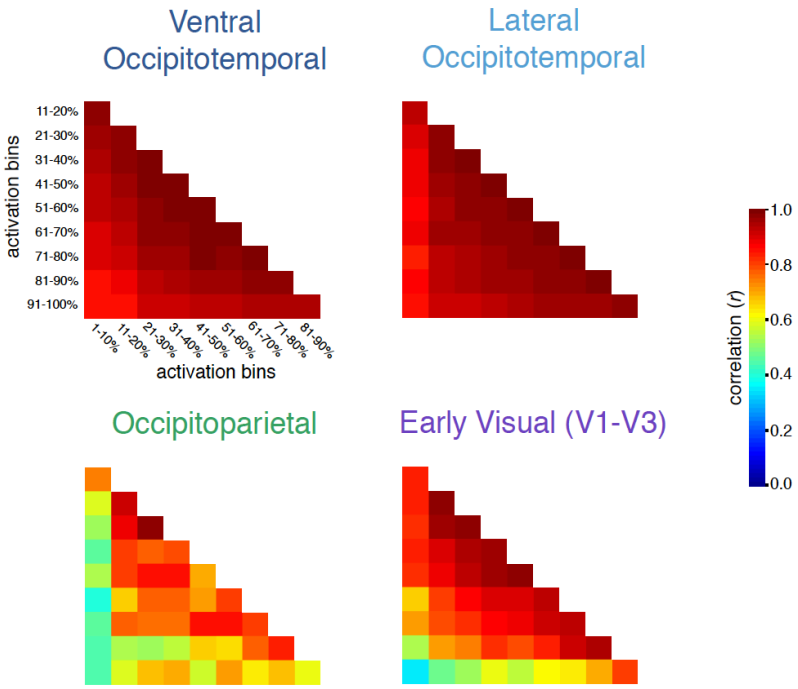

Figure C.5: Correlations of the similarity structures between each of the 10% activation bins in each of the four macro-scale sectors. Warmer colors denote higher correlations/similarity, while cooler colors denote lower correlations/similarity. Note that the correlations between each activation bin in ventral and occipitotemporal cortex are statistically significant.

References

Afraz, S.R., Kiani, R., & Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. *Nature,* 442, 692-695.

Alais, D., & Blake, R. (2005) *Binocular Rivalry*. Cambridge: MIT Press.

Alvarez, G.A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science,* 15, 106–111.

Alvarez, G. A., Horowitz, T. S., Arsenio, H. C., DiMase, J. S., & Wolfe, J. M. (2005). Do multielement visual tracking and visual search draw continuously on the same visual attention resources? *Journal of Experimental Psychology: Human Perception and Performance,* 31, 643-647.

Awh, E., Serences, J., Laurey, P., Dhaliwal, H., van der Jagt, T., & Dassonville, P. (2004). Unimpaired face discrimination during the attentional blink: Evidence for multiple processing channels. *Cognitive Psychology*, 48, 95-126.

Awh, E., Barton, B., & Vogel, E.K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science,* 18, 622-628.

Baayen, R.H. (2009). language: Data sets and function with "Analyzing linguistic Data: A practical introduction to statistics." R package version 0.955.

Baddeley, A.D., & Hitch, G. (1974). Working memory. In *The psychology of learning and motivation: Advances in research and theory*. ed. G.H. Bower (Academic Press: New York) pp 47–89.

Baldassi, C., Alemi-Neissi, A., Pagan, M., DiCarlo, J.J., Zecchina, R., & Zoccolan, D. (2013). Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons. *PLoS Computational Biology*, 9, e1003167.

Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255-278.

Bates, D.M., & Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-32.

Beck, D.M., & Kastner, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nature Neuroscience,* 8, 1110–1116.

Bell, A.H., Hadj-Bouziane, F., Frihauf, J.B., Tootell, R.B., & Ungerleider, L.G. (2009). Object representations in the temporal cortex of monkeys and humans as revealed by functional magnetic resonance imaging. *Journal of Neurophysiology*, 101, 688-700.

Bell, A., Malecek, N.J., Morin, E.L., Hadj-Bouziane, F., Tootell, R.B.H., & Ungerleider, L.G. (2011). Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity. *Journal of Neuroscience,* 31, 12229-12240.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review,* 94, 115–147.

Block, N. (2005). Two neural correlates of consciousness. *Trends in Cognitive Sciences*, 9, 46–52.

Brainard D.H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433-436.

Breitmeyer, B.G. (2007). Visual masking: Past accomplishments, present status, future developments. *Advances in Cognitive Psychology*, 3, 9-20.

Broadbent, D. (1958). *Perception and Communication*. London: Pergamon Press.

Caramazza, A., & Shelton, J.R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, 10, 1-34.

Carandini, M., Movshon, J.A., & Ferster, D. (1998). Pattern adaptation and cross-orientation interactions in the primary visual cortex. *Neuropharmacology,* 37, 501-511.

Carlson, T.A., Ritchie, J.B., Kriegeskorte, N., Durvasula, S., & Ma, J. (2013). Reaction time for object categorization is predicted by representational distance. *Journal of Cognitive Neuroscience.* 26, 132-142.

Cavanagh, P. (2011). Visual cognition. *Vision Research*, 13, 1538-1551.

Chao, L.L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage*, 12, 478-484.

Chao, L.L., Weisberg, J., & Martin, A. (2002). Experience-dependent modulation of category related cortical activity. *Cereberal Cortex*, 12, 545-51.

Chun, M.M., & Wolfe, J.M. (2001). Visual Attention. In E.B. Goldstein (Ed.) *Blackwell's Handbook of Perception* (pp. 272-310) Oxford: Blackwell.

Chklovskii, D.B., & Koulakov, A.A. (2004) Maps in the brain: What can we learn from them? *Annual Review of Neuroscience,* 27, 369-392.

Cichy, R.M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, 17, 455-464.

Cohen, M.A., Alvarez, G.A., & Nakayama, K. (2011). Natural-scene perception requires attention. *Psychological Science*, 22, 1165-1172.

Cohen, M.A., & Dennett, D.C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, 15, 358-364.

Cohen, M.A., Cavanagh, P., Chun, M.M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in Cognitive Sciences*, 16, 411-417.

Cox, D.D. (2014). Do we understand high-level vision? *Current Opinion of Neurobiology*, 25, 187-193.

Çukur, T., Nishimoto, T., Huth, A.G., & Gallant, J.L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, 16, 763-770.

Curtis, C.I., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences,* 7, 415-423.

Dehaene, S., & Changeux, J.P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70, 200–227.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, 15, 254-262.

Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology*, 5, e260.

Delvenne, J-F. (2005). The capacity of visual short-term memory within and between hemifields. *Cognition*, 96, B79-B88.

Delvenne, J-F., & Holt, J.L. (2012). Splitting attention across the two visual fields in visual short-term memory. *Cognition*, 12, 258-263.

Dennett, D.C. (2005). *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness.* Cambridge: MIT Press.

Desimone R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222.

DiCarlo, J.J., & Cox, D.D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, 11, 333-341.

DiCarlo, J.J., Zoccolan, D., & Rust, N.C. (2012). How does the brain solve visual object recognition? *Neuron*, 73, 415-434.

Dilks, D.D., Julian, J.B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area (OPA) is causally and selectively involved in scene perception. *Journal of Neuroscience*, *33*, 1331-1336.

Di Lollo, V., Enns, J.T., & Rensink, R.A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General*, 129, 481-507.

Downing, P.E., Jiang, Y., Shumann, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293, 2470-2473.

Drew, T., & Vogel, E.K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *Journal of Neuroscience,* 28, 4183-4191.

Duncan, J., & Humphreys, G.W. (1989). Visual search and stimulus similarity. *Psychological Review,* 96, 433-458.

Duncan, J., Martens, S., & Ward R (1997). Restricted attentional capacity within but not between sensory modalities. *Nature,* 387, 808-810.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 382, 598-601.

Epstein, R., DeYoe, E.A., Press, D.Z., Rosen, A.C., & Kanwisher, N. (2001). Neuropsychological evidence for a topographical learning mechanism in parahippocampal cortex. *Cognitive Neuropsychology*, 18, 481-508.

Eramudugolla, R., Irvine, D., McAnally, K.I., Martin, R.L., & Mattingly, J.B. (2005) Directed attention eliminates change deafness in complex auditory scenes. *Current Biology*, 15, 1008-1113.

Farah, M. (2004). Visual Agnosia. (2nd ed.) Cambridge: MIT Press.

Franconeri, S.L., Alvarez, G.A., & Enns, JT (2007). How many locations can you select? *Journal of Experimental Psychology: Human Perception and Performance,* 33, 1003-1012.

Franconeri, S.L., Jonathan, S.V., & Scimeca, J.M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science,* 21, 920-925.

Franconeri, S.L., Alvarez, G.A., & Cavanagh, P. (2013). Flexible cognitive resources: competitive content maps for attention and memory. *Trends in Cognitive Sciences*, 17, 134-141.

Freeman, J., Ziemba, C.M., Heeger, D.J., Simoncelli, E.P., & Movshon, J.A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience*, 16, 974-981.

Fedorenko, E., McDermott, J., Norman-Haignere, S., & Kanwisher, N. (2012). Sensitivity to musical structure in the human brain. *Journal of Neurophysiology*, 108, 3289-3300.

Gauthier, I., Skudlarski, P., Gore, J.C., & Anderson, A.W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience,* 3, 191-197.

Gazzaniga, M., Ivry, R.B., & Mangun, G.R. (2008). *Cognitive Neuroscience*. New York: W.W. Nortan & Company.

Goodale, M.A., & Milner, A.D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences* 15, 20-25.

Harel, A., Kravitz, D.J., & Baker, C.I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences U.S.A.,* 11, 962-971.

Haushofer, J., Livingstone, M.S., & Kanwisher, N. (2008). Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. *PLoS Biology*, 6, e187.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., & Pietrini P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science,* 293, 2425–2430.

Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke, M., & Ramadge, P.J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72, 404-416.

Hellige, J.B., Walsh, D.A., Lawrence, V.W., & Prasse, M. (1979). Figural relationship effects and mechanisms of visual masking. *Journal of Experimental Psychology: Human Perception and Performance,* 5, 88-100.

Hershler, O., Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research*, 45, 1707-1724.

Huth, A.G., Nishimoto, S., Vu, A.T., & Gallant, J.L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human cortex. *Neuron*, 76, 1210-1224.

Jackendoff, R. (1987). Consciousness and the Computational Mind, MIT Press.

Kanwisher, N. (2010). Functional specificity in the human brain: a window in- to the functional architecture of the mind. *Proceedings of the National Academy of Sciences U.S.A.,* 107, 11163-11170.

Kass, J.H. (1989). Why does the brain have so many visual areas? *Journal of Cognitive Neuroscience*, 1, 121-135.

Kastner, S., De Weerd, P., Pinsk, M.A., Elizondo, M.I., Desimone, R., & Ungerleider, R.G. (2001). Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. *Journal of Neurophysiology,* 86, 1398-1411.

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L.G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional fMRI. *Science*, 282, 108-111.

Keller, L.F., Arcese, P., Smith, J.N., Hochachka, W.M., & Stearns, S.C. (1994). Selection against inbred song sparrows during a natural population bottleneck. *Nature*, 372, 356-357.

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, 97, 4296-4309.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences*, 11, 16–22.

Konen, C.S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience*, 11, 224-231.

Konen, C.S., Behrmann, M., Nishimura, M., & Kastner, S. (2011). The functional neuroanatomy of visual agnosia: A case study. *Neuron*, 71, 49-60.

Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipito-temporal cortex. *Neuron*, 74, 1114-1124.

Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *Journal of Neuroscience*, 33, 10235-10242.

Kornblish, S., Cheng, X., Ohayon, S., & Tsao, D.Y. (2013). A network for scene processing in the macaque temporal lobe. *Neuron*, 79, 766-781.

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review. *Philosophical Transactions of the Royal Society of London B*, 362, 857-875.

Kourtzi, Z., & Connor, C.E. (2011). Neural representations for object perception: structure, category, and adaptive coding. *Annual Review of Neuroscience*, 34, 45-67.

Kravitz, D.J., Vinson, L.D., & Baker, C.I. (2008). How position dependent is visual object recognition? *Trends in Cognitive Sciences,* 12, 114–122.

Kravitz, D.J., Saleem, K.S., Baker, C.I., Ungerleider, L.G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, 17, 26-49.

Kriegeskorte N, Goebel, R, Bandettini P (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences U.S.A.*, 103, 3863-3868.

Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., Pandettini, P. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60, 1126-1141.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008b). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience,* 2, 1-28.

Kriegeskorte, N., & Kievit, R.A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences,* 17, 401-412.

Kovacs, G., Vogels, R., & Organ, G.A. (1995). Cortical correlate of pattern backward masking. *Proceedings of the National Academy of Sciences, USA*, 92, 5587-5591.

Lamme, V.A.F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, 1, 204–220.

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of Optical Society of America*, 70, 1458-147l.

Lehky, S.R., & Sereno, A.B. (2007). Comparison of shape encoding in primate dorsal and ventral visual pathways. *Journal of Neurophysiology*, 97, 307-319.

Li, F.F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, U.S.A., 99, 9596–9601.

Loftus, G.R., & Masson, M.E.J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin and Review,* 1, 476-490.

Ma, W.J., Husain, M., & Bays, P.M. (2014). Changing concepts of working memory. *Nature Neuroscience,* 17, 347-356

MacEvoy, S.P., Tucker, T.R., Fitzpatrick, D. (2009). A precide form of divisive suppression supports population coding in primary visual cortex. *Nature Neurosciences,* 12, 637-645.

Mahon, B.Z, & Caramazza, A. (2009). Concepts & Categories: A Cognitive Neuropsychological Perspective. *Annual Review of Psychology,* 60, 27–51.

Malinowski, P., & Hübner, R. (2001). The effect of familiarity on visual-search performance: Evidence for learned basic features. *Perception & Psychophysics*, *63*, 458-463.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Neuroscience*, 58, 25-45.

McKeeff, T.J., McGugin, R.W., Tong, F., & Gauthier, I. (2010). Expertise increases the functional overlap between face and object perception. *Cognition,* 117, 355-360.

McGugin, R.W., McKeeff, T.J., Tong, F., & Gauthier, I. (2011). Irrelevant objects of expertise compete with faces during visual search. *Attention Perception and Psychophysics,* 73, 309-317.

Michaels, C.F., & Turvey, M.T. (1979). Central sources o f visual masking: Indexing structure supporting seeing at a single brief glance. *Psychological Research*, 4, 1-61.

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience,* 6, 182-189

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-IT object representation. *Frontiers in Psychology*, 4, e128.

Nakayama, K., & Martini, P. (2011). Situating visual search. *Vision Research, 51, 1526-1537.*

Norman-Haignere, S., Kanwisher, N., & McDermott, J.H. (2013). Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *Journal of Neuroscience*, 33, 19451-19469.

Nunnally Jr, J. C. (1970). *Introduction to psychological measurement*. New York: McGraw-Hill.

Olshausen, B.A. & Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381, 607-609.

Olsson, H., & Poom, L. (2005). Visual memory needs categories. *Proceedings of the National Academy of Sciences*, U.S.A., 102, 8776-8780.

Op de Beeck, H.P., Haushofer, J., & Kanwisher, N. (2008). Interpreting fMRI data: Maps, modules, and dimensions. *Nature Reviews Neuroscience*, 9, 123-135.

O'Regan, J.K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. Behavioral and Brain Sciences, 24, 939–973.

O'Toole, A.J., Jiang, F., Abdi, H., & Haxby, J.V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, 17, 580-590.

Park, J., Carp, J., Hebrank, A., Park, D.C., & Polk, T.A. (2010). Neural specificity predicts fluid processing ability in older adults. *Journal of Neuroscience,* 30, 9253-9259.

Peelen M.V., Downing, P.E. (2005). Within-subject reproducibility of category-specific visual activation with functional MRI. *Human Brain Mapping,* 25, 402-408.

Peelen, M.V., & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences USA, 108,* 12125-12130.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Visio,n* 10, 437-442.

Pelli, D.G., & Tillman, K.A. (2008). The uncrowded window of object recognition. *Nature Neuroscience,* 11, 1129-1135.

Prinz, J. (2005). A neurofunctional theory of consciousness. In: Brook A and Akins K (eds.) *Cognition and the Brain: Philosophy and Neuroscience Movement* (pp. 381–396) Cambridge: Cambridge University Press.

Pitcher, D., Charles, L., Devlin, J.T., Walsh, V., & Duchaine, B. (2009) Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Current Biology*, 19, 319-324.

Pylyshyn, Z., & Storm, R.W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision,* 3, 179–197.

R Development Core Team (2008). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Reddy, L., & Kanwisher, N. (2007). Category selectivity in the ventral visual pathway confers robustness to cluster and diverted attention. *Current Biology,* 17, 2067-2072.

Reynolds, J.H., & Deismone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron,* 24, 19-29.

Romero, M.C., Pani, P., & Janssen, P. (2014). Coding of shape features in the Macaque anterior intraparietal area. *Journal of Neuroscience*, 34, 4006-4021.

Sato, T., Uchida, G., Lescroart, M.D., Kitazono, J., Okada, M., & Tanifuji, M. (2013). Object representation in inferior temporal cortex is organized hierarchically in a mosaic-like structure. *Journal of Neuroscience*, 33, 16642-16656.

Scalf, P.E., Torralbo, A., Tapia, E. & Beck, D.M. (2013). Competition explains limited attention and perceptual resources: implications for perceptual load and dilution theories. *Frontiers in Psychology,* 4, 243.

Sereno, A.B., & Maunsell, J.H.R. (1998). Shape selectivities in primate lateral intraparietal cortex. *Nature,* 395, 500-503.

Sereno M.I., Dale A.M., Reppas J.B., Dwong K.K., Belliveau J.W., Brady T.J., Rosen B.R., & Tootell R.B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268, 889-893.

Sereno, M.E., Trinath, T., Augath, M., & Logothetis, N.K. (2002). Three-dimensional shape representation in monkey cortex. *Neuron,* 33, 635-652.

Sinha, P., Balas, B.J., Ostrovsky, Y., & Russell, R. (2006). Face recognition by humans: 19 results all computer vision researchers should know about. *Proceedings of the IEEE.* 94, 1948-1962.

Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron,* 35, 1157-1165.

Sripati, A.P., & Olson, C.R. (2010). Global image dissimilarity in macaque inferotemporal cortex predicts human visual search efficiency. *Journal of Neuroscience*, 30, 1258-1269.

Suzuki, S., & Cavanagh, P. (1995). Facial organization blocks access to low-level features: An object inferiority effect. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 901-909.

Tanaka, K. (2003). Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cereberal Cortex*, 13, 1990-1999.

Tarr, M.J., & Bülthoff, H.H. (1998). Image-based object recognition in man, monkey, and machine. *Cognition,* 67, 1-20.

Tarr, M.J. & Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience,* 3, 764-769.

Taylor, J., & Downing, P. (2011). Division of labor between lateral and ventral extrastriate representations of faces, bodies, and objects. *Journal of Cognitive Neuroscience*, 23, 4122-4137.

Tong, F., & Pratte, M.S. (2012). Decoding patterns of human brain activity. *Annual Review of Psychology*, 63, 483-509.

Tsao, D.Y., Friewald, W.A., Knutsen, T.A., Mandeville, J.B., & Tootell, R.B.H. (2003). The representation of faces and objects in Macaque cerebral cortex. *Nature Neuroscience*, 6, 989-995.

Tsao, D.Y., Freiwald, W.A., Tootell, R.B.H., & Livingstone, M.S.L. (2006). A cortical region consisting entirely of face cells. *Science*, 311, 670-674.

Treue, S., Hol, K. & Rauber, H-J. (2000). Seeing multiple directions of motion—physiology and psychophysics. *Nature Neuroscience,* 3, 270-276.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327.

Ullman, S. (2006). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences,* 11, 58-64.

Ungerleider, L.G. & Bell, A.H. (2011). Uncovering the visual "alphabet": Advances in our understanding of object perception. *Vision Research*, 51, 782-799.

Ungerleider, L.G., & Mishkin, M. (1982). Two cortical visual systems. In *Analysis of visual behavior.* eds. D.J. Ingle, M.A. Goodale, R.J.W. Mansifield (MIT Press, Cambridge) pp. 549-583.

Wandell, B.A. (1999). Computational neuroimaging of human visual cortex. *Annual Review of Neuroscience,* 22, 145-173.

Wandell, B.A., & Dumoulin, S.O., & Brewer, A.A. (2007). Visual field maps in human cortex. *Neuron,* 56, 366-393.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception & psychophysics*, *56*, 495-500.

Watson A.B, & Pelli, D.G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception and Psychophysics*, 33, 113–120.

Wickens, C.D. (2008). Multiple resources and mental workload. *Human Factors*, 50, 449-455.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods,* 42, 671–684.

Williams, E.J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Scientific Research*, 2, 149-168.

Williams, M., Dang, S., & Kanwisher, N. (2007). Only some spatial patterns of fMRI response are read out in task performance. *Nature Neuroscience*, 10, 685-6863

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv: 1305.

Wolfe, J.M., & Horowitz, T.S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5, 1-7.

Wong, J.H., Peterson, M.S., & Thompson, J.C. (2008). Visual working memory capacity for objects from different categories: A face-specific maintenance effect. *Cognition,* 108, 719-731.

Xu, Y., & Chun, M.M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, 440, 91-95.

Zatorre, R.J., Belin, P., & Penhune, V.B. (2002) Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, 6, 37-46.

Zhang, W., & Luck, S. (2008). Discrete fixed-resolution representations in visual working memory. *Nature,* 453, 233–235.